

UNIVERZA V LJUBLJANI
FILOZOFSKA FAKULTETA
ODDELEK ZA PREVAJALSTVO

ŠPELA REHER

**Slovenščina na prepihu: kodno preklapljanje v objavah
slovenskih uporabnikov Twitterja**

Kvalitativna in kvantitativna analiza tвитov iz korpusa nestandardne slovenščine Janes

Magistrsko delo

Ljubljana, 2017

UNIVERZA V LJUBLJANI
FILOZOFSKA FAKULTETA
ODDELEK ZA PREVAJALSTVO

ŠPELA REHER

**Slovenščina na prepihu: kodno preklapljanje v objavah
slovenskih uporabnikov Twitterja**

Kvantitativna in kvalitativna analiza tvitov iz korpusa nestandardne slovenščine Janes

Magistrsko delo

Mentorica: doc. dr. Darja Fišer

Študijski program: Prevajanje
(slovenščina/angleščina/nemščina)

Ljubljana, 2017

Zahvale

Zahvaljujem se ...

- ... mentorici Darji Fišer za vse raziskovalne ideje in strokovno usmerjanje pri načrtovanju raziskave in pisanju magistrskega dela,
- ... Tomažu Erjavcu z Inštituta Jožef Stefan za tehnično podporo pri ustvarjanju korpusa preklpov in izvozu podatkov,
- ... Jaki Čibeju in Vanji Novak za prijateljsko pomoč, izmenjane zamisli in dragocene predloge v vseh fazah nastajanja tega dela,
- ... staršem in sestri za podporo in spodbudne besede, kadar sem jih potrebovala,
- ... Matjažu za razumevanje, oporo in potrpežljivost v dolgih delovnih dneh, ki jim včasih ni bilo videti konca.

Izvleček

Kodno preklapljanje – uporaba besed iz dveh ali več jezikov znotraj enega stavka oziroma diskurza – je ena od značilnosti računalniško posredovane komunikacije, še zlasti na večjezičnem internetu. Večina dosedanjih raziskav preklapljanja je bila osredotočenih na govorni diskurz, v slovenskem jeziku pa ta pojav do sedaj še ni bil obsežno raziskan. Cilj tega dela je preučiti kodno preklapljanje v tvitih, zbranih v projektu Janes – *Jezikoslovna analiza nestandardne slovenščine*. V teoretičnih izhodiščih so predstavljene značilnosti jezika na internetu in družbenih omrežjih (še posebej na Twitterju), jeziki v stiku, sorodne raziskave o kodnem preklapljanju in korpus Janes. Raziskovalni del sestavljata kvantitativna in kvalitativna analiza, zanj pa sta bila pripravljena dva vzorca po 1600 tvitov, ki so bili označeni v programu WebAnno na petih ravneh: jezik (angleščina, nemščina itd.), vrsta preklopa (medstavčni in znotrajstavčni), ortografska integracija (podomačen, delno podomačen ali nepodomačen zapis), morfologija (z razvidno končnico in/ali obrazilom ali brez) ter besedna vrsta (samostalniška zveza, medmet itd.). Rezultati so pokazali, da preklapljanje v slovenskih tvitih ni redek pojav. Delež znotrajstavčnih prekopov je višji od medstavčnih (2 : 1). V povprečju so kodni preklopi vsebovali 1,26 besed, približno 50 % vseh prekopov pa je bilo enobesednih. Med zaznanimi jeziki prevladuje angleščina (90 %), ki ji sledita srbohrvaščina (4,6 %) in nemščina (3,3 %), skupno pa smo v označenem vzorcu zaznali devet tujih jezikov. Večina kodnih prekopov ni bila ortografsko ali morfološko integriranih; samo pri germanizmih je bila stopnja integracije višja. Iz podatkov lahko sklepamo, da v nasprotju z nekaterimi trditvami v sorodnih raziskavah preklopi med slovenščino in angleščino zajemajo tudi zaprte besedne vrste, ne le odprti, poleg tega ni bilo pomembnih razlik v preklapljanju med moškimi in ženskimi avtorji. Kvalitativna analiza je pokazala, da kodno preklapljanje opravlja številne funkcije, kot so referenčna, ekspresivna in poudarjalna, ter prispeva h gospodarni rabi jezika. V zvezi s semantičnimi polji smo ugotovili, da je kodno preklapljanje pogosto povezano s popularno kulturo, zlasti TV-oddajami, športom, (kulturno specifično) hrano in Twitterjem. Mnogi kodni preklopi so bili idiomi, frazni glagoli, kolokacije ali pregovori. Na podlagi raziskave lahko zaključimo, da kodno preklapljanje ne služi zgolj zapolnjevanju leksikalnih vrzeli, temveč raznolikim namenom, kar kaže tudi na to, da vsaj nekateri uporabniki Twitterja dobro poznajo tuje jezike, predvsem angleščino, ter to izkoriščajo za čim boljše izražanje zelenega sporočila.

Ključne besede: kodno preklapljanje, prevzete besede, računalniško posredovana komunikacija, korpusno jezikoslovje, slovnične omejitve, diskurzne funkcije

Abstract

Code-switching (hereafter CS) – the use of words from two or more languages within one sentence or utterance – is one of the characteristics of computer-mediated communication, especially on the multilingual internet. The majority of previous studies on CS focused on spoken discourse, and there have been only a few studies on CS in the Slovenian language until now. Therefore, this thesis aims to explore CS in the tweets collected as part of the Janes corpus of internet Slovene. The theoretical background introduces the language on the internet and on the social media (on Twitter in particular), languages in contact, previous research on code-switching and the Janes corpus. The empirical part is divided into the quantitative and qualitative analyses, for which two samples of 1600 tweets were created and annotated in WebAnno on five levels: language (English, German etc.), type of CS (intra- and intersentential), orthography (assimilated, partially assimilated or non-assimilated), morphology (not evident or evident from the ending/affix) and part of speech (noun phrase, verb phrase etc.). The results showed that CS is not a rare phenomenon in the tweets written by Slovenian users. Intrasentential CS is more common than intersentential (2 : 1). On average, code-switches consisted of 1.26 words and about 50% of them were single-word switches. Among the languages found in the corpus, English clearly dominates (90%), followed by Serbo-Croatian (4.6%) and German (3.3%). In total, nine different foreign languages were detected in the annotated sample. Most code-switches were not orthographically or morphologically integrated; only German items were assimilated to a greater degree. According to the results, the constraint on the switchability of closed-class words proposed by some researchers does not apply in Slovenian-English CS and no significant differences in CS between male and female authors were found. The qualitative analysis showed that CS fulfils various functions, such as referential, expressive, or phatic, and contributes to language economy. In terms of the semantic fields, CS was related to popular culture, in particular TV shows, sport, (culturally specific) food and Twitter. Many code-switches were idiomatic expressions, phrasal verbs, collocations and even proverbs. To conclude, the research showed that CS is not used only to fill in the lexical gaps but for a variety of purposes, which shows that at least some Twitter users are quite competent in foreign languages, especially English, and can use them to convey their message in the best possible way.

Keywords: code-switching, borrowing, computer mediated communication, corpus linguistics, grammatical constraints, discourse functions

Kazalo

1	Uvod.....	1
1.1	Zgradba magistrskega dela	1
1.2	Namen in hipoteze	2
2	Teoretična izhodišča	4
2.1	Jezik na internetu	4
2.1.1	Splošne značilnosti	5
2.1.2	Vpliv tehnologije na jezik na internetu	7
2.2	Jezik na družbenih omrežjih (jezik uporabniško generiranih vsebin, splet 2.0).....	9
2.2.1	Raziskave jezikovne rabe na Twitterju	11
2.3	Jeziki v stiku	13
2.3.1	Vpliv angleščine na druge jezike	14
2.4	Kodno preklapljanje	17
2.4.1	Razmejitev od kodnega mešanja in prevzetih besed	18
2.4.2	Vrste kodnega preklapljanja	20
2.4.2.1	Situacijsko in metaforično preklapljanje.....	20
2.4.2.2	Znotrajstavčno in medstavčno preklapljanje.....	20
2.4.3	Dosedanje raziskave	22
2.4.4	Kodno preklapljanje v računalniško posredovani komunikaciji	23
2.4.5	Okvir matričnega jezika (Matrix Language Framework).....	25
2.4.6	Slovnične omejitve pri kodnem preklapljanju	26
2.4.6.1	Omejitev glede strukturne integritete (equivalence constraint, structural-integrity constraint)	26
2.4.6.2	Omejitev glede prostih morfemov (free-morpheme constraint)	27
2.4.6.3	Omejitev glede možnosti preklapljanja zaprtih besednih vrst (constraint on switchability of closed-class groups)	27
2.4.6.4	Hierarhija otokov vrinjenega jezika	27
2.4.7	Funkcije kodnega preklapljanja	28
2.5	Korpus Janes	31
2.5.1	Podkorpus tvitov	32
3	Raziskovalni del.....	33

3.1	Načrt raziskave	33
3.1.1	Kvantitativna analiza.....	34
3.1.2	Kvalitativna analiza.....	37
3.2	Priprava in označevanje korpusa preklapov	39
3.2.1	Priprava korpusa.....	39
3.2.2	Označevanje preklapov	40
3.3	Kvantitativna analiza	45
3.3.1	Splošne značilnosti preklapljanja	45
3.3.1.1	Pogostost preklapov	45
3.3.1.2	Uporabljeni jeziki.....	47
3.3.1.3	Dolžina preklapov	49
3.3.2	Slovnične omejitve preklapljanja	52
3.3.3	Integracija preklapov v slovenščino.....	55
3.3.3.1	Ortografska integracija (zapis).....	55
3.3.3.2	Morfološka integracija	57
3.3.3.3	Ortografska in morfološka integracija	59
3.3.4	Razlike glede na spol.....	59
3.4	Kvalitativna analiza	61
3.4.1	Diskurzne funkcije kodnega preklapljanja	61
3.4.1.1	Zgoščeno izražanje (gospodarna raba jezika)	61
3.4.1.2	Ponovitve	61
3.4.1.3	Odzivanje oz. navezovanje	62
3.4.1.4	(Retorična) vprašanja	62
3.4.1.5	Odgovor sogovorniku	62
3.4.1.6	Vljudnostni izrazi.....	63
3.4.1.7	Izražanje čustev	63
3.4.1.8	Medmeti – klepetalniške krajšave.....	63
3.4.1.9	Ekspresivni elementi	64
3.4.2	Tematska polja	64
3.4.2.1	Znani citati	64
3.4.2.2	Televizijske serije, filmi.....	64
3.4.2.3	IT, internet.....	65

3.4.2.4	Šport.....	65
3.4.2.5	Kulturno specifični pojmi	65
3.4.2.6	Twitter.....	66
3.4.3	Frazeologija.....	66
3.4.3.1	Pregovori.....	66
3.4.3.2	Idiomi	67
3.4.3.3	Frazni glagoli	67
3.4.3.4	Stalne besedne zveze in kolokacije	67
4	Zaključek	68
5	Summary.....	72
6	Viri in literatura	76
7	Priloge.....	83
7.1	Priloga A: Smernice za označevanje kodnega preklapljanja v korpusu slovenskih tvitov Janes	83
7.2	Priloga B: Dodatni primeri zanimivih kodnih preklapov	90
7.3	Priloga C: Seznami kodnih preklapov po besednih vrstah.....	91
7.4	Priloga D: Seznami kodnih preklapov glede na integracijo	93
7.5	Priloga E: Seznam najpogostejših besed, ki so del kodnih preklapov	96
8	Izjava o avtorstvu.....	97

Kazalo slik

Slika 1: Časovnica spleta 2.0 z izbranimi spletnimi stranmi (povzeto po Herring 2012: 3).....	9
Slika 2: Načrtovan potek raziskave.....	33
Slika 3: Značilnosti tvitov, vključenih v prvi raziskovalni vzorec.....	39
Slika 4: Značilnosti tvitov, vključenih v drugi raziskovalni vzorec.....	40
Slika 5: Shema poteka označevanja kodnih preklapov v programu WebAnno.	41
Slika 6: Posnetek zaslona s petimi tviti po končanem označevanju.	44

Kazalo tabel

Tabela 1: Število in delež tvitov z vsaj enim kodnim preklpom.....	45
Tabela 2: Število in delež preklapov v korpusu preklapov.....	46
Tabela 3: Število in delež heštegov v kodnih preklpkih.	47

Tabela 4: Število in delež preklpov po jezikih v korpusu preklpov.....	47
Tabela 5: Deleži jezikov glede na vrsto preklpa.	48
Tabela 6: Deleži preklpov glede na dolžino in vrsto preklpa.....	49
Tabela 7: Povprečna dolžina preklpov glede na vrsto preklpa.....	50
Tabela 8: Deleži preklpov glede na dolžino in jezik preklpa.	50
Tabela 9: Povprečna dolžina preklpov glede na jezik.....	51
Tabela 10: Deleži preklpov glede na dolžino in besedno vrsto.....	52
Tabela 11: Deleži preklpov glede na besedno vrsto za tri najpogostejše jezike.	54
Tabela 12: Ortografska integracija glede na besedne vrste.....	56
Tabela 13: Morfološka integracija glede na besedne vrste.	58
Tabela 14: Pogostost preklpov v tvitih ženskih in moških avtorjev.	60
Tabela 15: Deleži jezikov v tvitih ženskih in moških avtorjev.	60
Tabela 16: Dolžina preklpov v tvitih ženskih in moških avtorjev.	60

Kazalo grafov

Graf 1: Prisotnost jezikov na spletu glede na število internetnih uporabnikov (Internet World Stats 2017).....	15
Graf 2: Razmerje jezikov na spletu glede na količino vsebin (W3Techs 2017).	15
Graf 3: Deleži preklpov glede na standardnost besedila v splošnem vzorcu (N = 507).	46
Graf 4: Deleži medstavčnih in znotrajstavčnih preklpov v posameznem jeziku.	49
Graf 5: Deleži preklpov glede na besedno vrsto.	53
Graf 6: Deleži preklpov glede na ortografsko integracijo.....	55
Graf 7: Deleži angleških in nemških preklpov glede na ortografsko integracijo.....	56
Graf 8: Deleži preklpov glede na morfološko integracijo.....	57
Graf 9: Deleži angleških in nemških preklpov glede na morfološko integracijo.....	58

1 Uvod

Internet je povezal ljudi z vseh koncev sveta in korenito spremenil marsikateri vidik vsakdanjega življenja. Za področje jezikoslovja med drugim pomeni prostor, v katerem prihaja do stikov med jeziki v obsegu, večjem kot kadar koli doslej. To ima pomemben vpliv tudi na rabo jezika, ki ga uporabniki prilagajajo novim oblikam sporočanja in možnostim izražanja. Lažje in pogostejše smo v stiku s tujimi jeziki, predvsem angleščino kot dominantnim jezikom interneta, zaradi česar hitreje prevzemamo tuje prvine – kar za nekatere pomeni bogatenje besedišča, pa za druge predstavlja perečo grožnjo, predvsem za manjše jezike.

Ena izmed posledic jezikovnih stikov je kodno preklapljanje, pri katerem ljudje iz različnih razlogov mešajo besede iz dveh ali več jezikov znotraj istega stavka ali diskurza. Predvsem v nestandardni komunikaciji mešanje jezikov ni redek pojav, kar lahko opazimo že med naključnim brskanjem po forumih, komentarjih na novice ali družbenih omrežjih, če bomo pozorni, pa se bomo tudi med klepeti s prijatelji hitro zalotili, kako nam kdaj prej pade na pamet tuja beseda kot slovenska. Situacije, v katerih preklapljamo, so mnoge in raznolike, zato smo za namen te magistrske naloge zožili polje raziskovanja, in sicer smo se osredotočili na kodno preklapljanje v tvitih slovenskih uporabnikov, ki so vključeni v korpus Janes.

1.1 Zgradba magistrskega dela

Magistrsko delo je razdeljeno na dva glavna dela, teoretičnega in raziskovalnega. Najprej bomo predstavili splošne značilnosti jezika na internetu in vpliv tehnologije na jezikovno rabo, nato pa se bomo v naslednjem poglavju podrobneje posvetili jeziku na družbenih omrežjih in Twitterju. Sledi poglavje o jezikih v stiku, v katerem se bomo dotaknili vprašanja, kakšen je vpliv angleščine na druge jezike in slovenščino. Nato bomo predstavili dosedanje raziskave o kodnem preklapljanju, in sicer bomo podali njegovo opredelitev, našli različne vrste preklapljanja, opisali različne pristope k raziskovanju tega pojava ter pripravili pregled slovničnih omejitev, ki so jih predlagali različni raziskovalci, in funkcij kodnega preklapljanja oziroma razlogov, zakaj se ga ljudje poslužujejo v različnih situacijah. Na koncu teoretičnih izhodišč sta predstavljena korpus nestandardne slovenščine Janes in podkorpus tvtov.

V raziskovalnem delu bomo predstavili potek dela, pripravo raziskovalnega korpusa in analizo, ki je razdeljena na dva dela. V kvantitativnem delu bomo na podlagi zbranih rezultatov opazovali splošne značilnosti preklapljanja (pogostost, zastopanost jezikov in dolžino kodnih prekopov), morebitne slovnične omejitve (katerih besednih vrst ni mogoče preklapljati) ter ortografsko in morfološko integracijo prekopov v slovenščino. Na kratko bomo preverili tudi,

ali je na podlagi zbranih podatkov mogoče ugotoviti kakšne razlike v preklapljanju glede na spol uporabnika. V okviru kvalitativne analize si bomo ogledali diskurzne funkcije, tematska polja in frazeološke pojave, ki se pojavljajo v označenem korpusu tvitov. Na koncu sledi še povzetek celotnega dela v angleškem jeziku.

1.2 Namen in hipoteze

Namen magistrskega dela je raziskati kodno preklapljanje v tvitih slovenskih uporabnikov. Na podlagi korpusa spletne slovenščine Janes bomo pripravili raziskovalni korpus tvitov, ki vsebujejo tujejezične prvine. Preklope (tj. dele besedil v tujem jeziku) bomo nato ročno označili v programu WebAnno, tako da bomo vsakemu pripisali ustrezne oznake za vrsto preklopa, jezik, prilagojenost zapisa, morfološko integracijo in besedno vrsto. Zbrane podatke bomo izvozili v format .xlsx in jih obdelali v programu Excel, kjer bomo pripravili relevantne tabele in grafe.

Že med označevanjem omenjenih značilnosti si bomo beležili tudi opažanja za kvalitativni del raziskave, v katerem bomo izpostavili diskurzne funkcije, ki razkrivajo razloge za preklapljanje v tuj jezik, tematska polja, v katerem se pogosto pojavljajo preklopi, ter frazeološke besedne zveze (pregovore, idiome, kolokacije), ki si jih uporabniki »izposojajo« v izvirni, tuji obliki.

Izhodiščne hipoteze, na podlagi katerih smo zasnovali raziskavo, so naslednje (v oklepaju je številka razdelka v raziskovalnem delu, kjer so obravnavane):

A. Splošne značilnosti preklapljanja (razdelek 3.3.1)

- 1) V nestandardnih besedilih bo delež prekopov višji kot v standardnih.
- 2) Med preklopi bodo prevladovale besede oz. stavki v angleščini, vendar bodo prisotni tudi drugi jeziki (nemščina, španščina, jeziki nekdanje SFRJ).
- 3) Več bo znotrajstavčnih prekopov kot medstavčnih, znotrajstavčni preklopi pa bodo v povprečju krajši kot medstavčni.

B. Slovnične omejitve preklapljanja (razdelek 3.3.2)

- 4) Med preklopi bo več polnopomenskih (samostalnikov, glagolov, pridevnikov) kot slovničnih besed (predlogov, veznikov, zaimkov).

C. Integracija prekopov v slovenščino (razdelek 3.3.3)

- 5) Večina kodnih prekopov bo ortografsko neprilagojenih oz. zapisanih v skladu z izvirnim jezikom.
- 6) Pri večini prekopov ne bo razvidne morfološke integracije, vendar bodo glagoli pogosteje integrirani (spregani s slovensko končnico) kot samostalniki in pridevniki.

D. Razlike glede na spol (razdelek 3.3.4)

- 7) Pogostost in značilnosti kodnih preklpov se ne bodo bistveno razlikovale glede na spol avtorja.

E. Diskurzne funkcije kodnega preklapljanja (razdelek 3.4.1)

- 8) Kodni preklopi bodo zapolnjevali leksikalne vrzeli ali pa bodo uporabljeni za doseganje posebnega poudarka.

F. Tematska polja (razdelek 3.4.2)

- 9) Kodni preklopi bodo pogosto povezani z informacijsko tehnologijo, popularno kulturo (televizijske serije, filmi, glasba) in kulturno specifičnimi pojavi.

G. Frazeologija (razdelek 3.4.3)

- 10) Med večbesednimi kodnimi preklopi bodo tudi idiomi in stalne besedne zveze.

Ob tem moramo poudariti, da cilj raziskave ni sistematična primerjava oziroma ugotavljanje, kakšen je vpliv sociolingvističnih in lingvističnih dejavnikov na kodno preklapljanje na splošno, temveč ponuditi kar se da celovit opis kodnega preklapljanja v specifični situaciji, tj. med slovenskimi uporabniki na Twitterju. Na podlagi našega korpusa preklpov prav tako ne bo mogoče podati zaključkov o tem, kako pogosto se uporabniki odločijo za mešanje jezikov iz točno določenega razloga ali za katero tematiko je kodno preklapljanje najbolj značilno.

Kljub temu pa bo raziskava podala vpogled v dejansko rabo slovenščine na družbenih omrežjih, kjer uporabniki odkrivajo inovativne načine, kako na zase najprimernejši način izraziti želeno sporočilo – tudi s poseganjem po tujejezičnih prvinah. Hkrati lahko predstavljene metode služijo kot podlaga za nadaljnje raziskave kodnega preklapljanja v slovenskem jeziku, saj je na tem področju še mnogo neodgovorjenih vprašanj.

2 Teoretična izhodišča

2.1 Jezik na internetu

Razvoj računalništva, še posebej interneta, je imel izjemen vpliv na mnoga področja sodobnega življenja. Za vse, ki se ukvarjajo z jezikom, je prinesel predvsem eksplozijo novih vsebin, ki so sicer temeljile na tradicionalnih medijih (novice, članki ipd.), vendar so se na internetu razvijale v nove oblike. Nekatere od njih se tako le malo ločijo od tradicionalnih situacij pisne komunikacije, druge pa se približujejo tipu interakcije, značilne za govor (Kranjc 2003: 445).

Preučevanje jezika, besedil, diskurzov in družbenih interakcij v digitalnih medijih oziroma v računalniško posredovani komunikaciji (angl. *Computer-Mediated Communication* ali CMC) je od zgodnjih 90. let 20. stoletja zraslo v samostojno raziskovalno področje, kar dokazujejo številni zborniki, posvečeni tej temi, kot tudi objave v uveljavljenih jezikoslovnih revijah (Georgakopoulou in Spilioti 2016: 1). Med najpomembnejšimi raziskovalci in utemeljitelji tega področja je Susan Herring, urednica temeljne zbirke *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives* (1996), med vplivnejšimi pa so zagotovo tudi Naomi Baron (*From Alphabet to E-mail* (2000) in *Always on: Language in an Online and Mobile World* (2008)), David Crystal (*Language and the Internet* (2006), *Internet linguistics: a student guide* (2011)) in Jannis Androutsopoulos (med drugim urednik posebne izdaje revije *Journal of Sociolinguistics: Sociolinguistics and Computer-Mediated Communication* (2006)).

Glede na relativno kratko zgodovino interneta, razvejanost jezikoslovnih raziskav in bliskovite spremembe v digitalnih tehnologijah ne preseneča, da so raziskovalci bolj ali manj isti predmet proučevanja poimenovali z različnimi nazivi. Med najbolj znanimi je Crystalov *Netspeak*, zaslediti pa je mogoče tudi izraze, kot so internetni jezik, kibergovorica (angl. *cyberspeak*), elektronski diskurz, elektronski jezik, e-jezik, interaktiven pisni diskurz ali računalniško posredovana komunikacija (Crystal 2006 in 2011, Baron 2008: 11). Herring (2012: 6) je izpostavila, da se komunikacijske tehnologije selijo z računalnikov tudi na mobilne telefone ter televizijo, zato nekateri raziskovalci uporabljajo izraze, kot je diskurz »digitalnih medijev« ali »novih medijev«, ki pa so po njenem mnenju vendarle preširoki. Crystal (2011: 2) za raziskovalno področje, ki se ukvarja z vsemi manifestacijami jezika v elektronskih medijih, kot najustreznejšega predlaga termin internetno jezikoslovje.

Računalniško posredovana komunikacija pomeni pošiljanje in/ali sprejemanje sporočil v kateremkoli naravnem jeziku prek računalniške povezave, poleg tega pa lahko zajema tudi druge tehnologije za komuniciranje na daljavo, denimo pošiljanje SMS-ov ali elektronske pošte

preko telefonov (Baron 2008: 10). Pravzaprav pokriva cel spekter možnosti – od besedil, ki so podobna tradicionalnim (ne-spletnim) pisnim besedilom, do dialogov, ki spominjajo na govor, razlika pri vseh je le v načinu prenosa sporočila (ibid.: 11). Ugotovimo lahko torej, da je računalniško posredovana komunikacija nekoliko širši pojem od jezika na internetu, vendar se oba izraza oziroma področji v večji meri prekrivata, zato jih tudi v tej nalogi ne bomo striktno ločevali.

Jezikovne študije računalniško posredovane komunikacije »prvega vala« so pri opisovanju jezikovnih pojavov izhajale iz dihotomije med sinhrono in asinhrono komunikacijo (Androutsopoulos 2006: 420), drug pogost kriterij za klasifikacijo različnih oblik elektronske komunikacije pa je število tvorcev in naslovnikov, tj. ali je sporočilo namenjeno eni osebi (angl. »one-to-one«) ali večjemu občinstvu (angl. »one-to-many«) (Baron 2008: 14). Poleg tega se je večina raziskav na začetku ukvarjala z danes že preseženim vprašanjem, ali je računalniško posredovana komunikacija bolj podobna govorjenim ali pisnim besedilom (Baron 2003: 23). Od 90. let 20. stoletja je na primer veljalo, da spletno komunikacijo zaznamuje več značilnosti, ki so bile tradicionalno povezane z govorom: interaktivnost, neformalnost, igrivost in povezane skupnosti (Tagg in Seargant 2014: 166). David Crystal (2006: 44) pa je ugotovil, da ima internetni diskurz pravzaprav več skupnega s pisnimi kot govorjenimi besedili, pri čemer zagovarja stališče, da ga je treba obravnavati kot posebno jezikovno varianto/različico, ki ima svoje besedišče, grafologijo, slovnico in pogoje uporabe.

2.1.1 Splošne značilnosti

Jezik, ki se pojavlja na spletu, je »fragmentiran, poln tipografskih napak, pogosto brez ločil in včasih povsem nekoherenten, poleg tega pa se dotika praktično vseh tem, ki si jih je mogoče zamisliti« (Baron 2003: 4). Najpogostejše jezikovne značilnosti, ki so bile predmet raziskav, vključujejo emotikone, nekonvencionalno črkovanje, značilnosti govorjenega jezika, značilnosti dialektov, opolzkosti in kodno preklapljanje (Androutsopoulos 2006: 421). Slednjemu se bomo obširneje posvetili v ločenem poglavju, zato je v nadaljevanju predstavljen samo kratek pregled najpomembnejših lastnosti jezika na internetu. Na začetku velja opozoriti tudi na to, da so številni raziskovalci pri tem, ko so utemeljevali generalizirane zaključke o jeziku na internetu, pravzaprav opisovali značilnosti angleščine, in še to pogosto zgolj v enem žanru računalniško posredovane komunikacije (Danet in Herring 2007: 5), vendar pa so v raziskavah v drugih jezikih odkrili podobne pojave (ibid.: 12). Tudi v slovenskem prostoru je to področje v zadnjih letih pritegnilo mnoge jezikoslovce, ki so razvili nove vire in orodja ter prišli do zanimivih raziskovalnih rezultatov v povezavi s slovenščino na spletu (Fišer 2016: i).

Leta 2006 je Crystal opisal naslednje situacije, v katerih se pojavlja netspeak, pri čemer pa je že sam opozoril, da je vsakršen poskus klasifikacije kmalu zastarel, saj se zaradi tehnoloških sprememb in razvoja novih medijev pojavljajo vedno nove različice (Crystal 2006: 10–15):

- elektronska pošta,
- klepeti (sinhroni in asinhroni),
- virtualni svetovi (večuporabniške spletne igre, npr. D&D),
- svetovni splet (*World Wide Web*): enciklopedije, arhivi, rumene strani, oglasi, samozaložništvo, igre, novice, kreativno pisanje, sklepanje poslov itd.,
- *instant messaging* (pogovori v realnem času),
- blogi (po letu 2000; lahko osebni dnevniki, prispevki o splošnih temah, hobijih ... lahko imajo enega ali več avtorjev).

Kot zanimivost in za prikaz, kako hitro potekajo spremembe na internetu, lahko omenimo, da zadnjih dveh alinej v prvi izdaji leta 2001 še ni bilo, v monografiji *Internet linguistics* pa je avtor na seznam dodal tudi družbena omrežja (Crystal 2011: 11). Kljub temu da se torej raba jezika med različnimi platformami precej razlikuje, je Crystal na splošno o jeziku na internetu ugotovil, da ima značilnosti tako govorjenega kot zapisanega jezika, ob tem pa so se razvile tudi določene unikatne značilnosti, zaradi katerih ga je treba obravnavati kot povsem nov medij. Kot razliko v primerjavi z govorom izpostavlja pomanjkanje sinhronega odziva, uporabo emotikonov in možnost sodelovanja v več pogovorih naenkrat, v primerjavi s pisnimi besedili pa poudarja možnost ustvarjanja povezav, stalnega spreminjanja besedil in soavtorstva več oseb (ibid.: 21–30).

Med jezikovnimi značilnostmi računalniško posredovane komunikacije tudi Baron (2003: 20–21) poleg emotikonov navaja okrajšave in kratice, ki so sicer značilne tudi za tradicionalna pisna besedila, toda v tem kontekstu imajo poleg ekonomičnosti tudi družbeno motivacijo, saj kažejo pripadnost spletni skupnosti. Poleg že omenjenih značilnosti so tipične ortografske značilnosti angleščine tudi krajšanje besed (npr. *gd* za *good*), homofoni (beseda/črka, npr. *u* za *you* ali *2* za *to*), stilizirano črkovanje (npr. *I'm soooooo happy*) ter nekonvencionalna/stilizirana raba loči (npr. *!!!!!!!*) (Page et al. 2014: 36–37). Tudi slovenski avtorji opažajo podobne spremembe, in sicer pogoste krajšave, emocionalno ikonografijo, spletne neologizme in neobičajno rabo ločil in simbolov (Michelizza 2007: 207) ali uporabo homofonov (Marko 2016: 48–53), pa tudi pisanje slovenskih črk č, š in ž brez strešic, rabo črk angleške abecede (w, y, x) in znakov (npr. &) ter podobno (Šabec 2014a: 403).

Jezik na internetu torej pokriva širok nabor možnosti – od pisnih izdelkov, ki se ne razlikujejo bistveno od nespletnih besedil (npr. akademski članki ali poslovna poročila), do komunikacije med osebami, ki je zelo podobna pogovorom v živo (Baron 2003: 11). V okviru ene spletne strani so lahko združene tudi različne oblike. Kot takšen primer lahko omenimo omrežje Facebook, kjer imajo uporabniki možnost klepetanja, pošiljanja sporočil (podobnih e-pošti), branja novic, spremljanja oglasov, ustvarjanja svojih vsebin itd. Pri proučevanju jezika na internetu se raziskovalci soočajo tudi s številnimi izzivi, kot so doslej neprimerljiv obseg podatkov, ki je na voljo na internetu, raznolikost jezikov in izjemna hitrost sprememb, zaradi katerih lahko vsaka študija interneta hitro postane zastarela (Crystal 2011: 10–11).

Dodaten prepih so v računalniško posredovano komunikacijo prinesla družbena omrežja, zato bomo njihov vpliv na jezik na internetu oz. jezik uporabniško generiranih besedil obravnavali v razdelku 2.2. Za razliko od prvih vplivnejših študij, ki so se osredotočale na en jezik, se je v zadnjih letih povečalo število raziskav medsebojnih vplivov različnih jezikov in pojavov, ki so rezultat stikov med jeziki. Ker gre za spoznanja, ključna za razumevanje kodnega preklapljanja, ki je osrednja tema tega diplomskega dela, se bomo jezikom v stiku podrobneje posvetili v razdelku 2.3, v nadaljevanju pa samo še nekaj besed o vplivu tehnologije na jezik na splošno.

2.1.2 Vpliv tehnologije na jezik na internetu

Naomi Baron je že leta 1984 objavila članek, v katerem je obravnavala vpliv računalniško posredovane komunikacije na spremembe v jeziku. Med drugim je ugotovila, da so udeleženci pogosto ločeni tako časovno kot krajevno; ta način ponuja več priložnosti za asinhrono komunikacijo, pri čemer so spletna sporočila trajnejša od govora, namesto parajezikovnih značilnosti, kot so kretnje in prozodija, pa uporabniki izkoriščajo vizualne možnosti (oblikovanje, simbole, slike); zaznamujejo jo tudi medjezikovne interakcije in mešanje različnih pisav (Tagg in Seargent 2014: 166).

Baron (2003: 4) kot pomembne dimenzije, ki jih je treba upoštevati pri preučevanju jezika na internetu, izpostavlja dihotomijo med govorom in pisanimi besedili, vpliv tehnologije ter medjezikovne in medkulturne vplive. Poleg tega izpostavlja štiri spremenljivke, ki lahko služijo za analizo posameznih oblik računalniško posredovane komunikacij, in sicer funkcijo sporočila, omejitve, ki izhajajo iz naprav (angl. *device constraints*), posebne jezikovne značilnosti in profil udeležencev. Ko govorimo o vplivu tehnologije, se pravzaprav dotikamo omejitev, ki izhajajo iz naprav. Tehnologija, ki se uporablja za prenos sporočila, lahko vpliva tako na obliko kot vsebino sporočila, npr. v primeru majhnih zaslonov in tipkovnic, na katerih je težje ustvariti ali brati daljša besedila (ibid.: 19–20).

Vpliv tehnološkega napredka na naravo in uporabo pisnega jezika ni nič novega. Izum tiskanja je denimo spodbudil standardizacijo črkovanja in rast pismenosti (Baron 2003: 7). Nova spletna okolja prinašajo nove vrste vsebin (posodobitve statusov, besedilni napisi na videoposnetkih, oznake itd.), kontekstov in občinstev. Nove komunikacijske tehnologije omogočajo besedilne klepete v množičnih večigralskih spletnih igrah, skupinsko urejanje dokumentov, sklepanje »prijateljstev« in zaprtih skupnosti (Facebook), družbeno označevanje/priporočanje. Vsa ta področja so že ali še bodo predmet različnih raziskav (Herring 2012: 6).

Napredek na področju računalništva je jezikoslovcem ponudil nova orodja za kvantitativne in kvalitativne analize, ki so spodbudila razvoj različnih disciplin in omogočila raziskovanje dotlej skritih vidikov jezika. Zato je vprašanje jezika na internetu povezano z računalniško obdelavo naravnega jezika (angl. *natural language processing* oz. NLP), ki je ključna za iskanje, pridobivanje in nadaljnjo uporabo informacij iz eksponentno naraščajočega obsega besedil na spletu. Predpostavka večine programov za računalniško obdelavo naravnega jezika je, da so nabor tem, slovnica in besedišče omejeni z določenimi omejitvami in konvencijami, toda raba jezika na internetu je vse prej kot konvencionalna ali zamejena (Baron 2003: 4). Za prilagajanje računalniških orodij in njihovo učinkovito uporabo pri obdelavi besedil in iskalnikov je torej treba razumeti jezikovne vzorce, ki se pojavljajo na spletu. To je pomembno tudi za razvoj drugih jezikovnih tehnologij, na primer strojnega prevajanja.

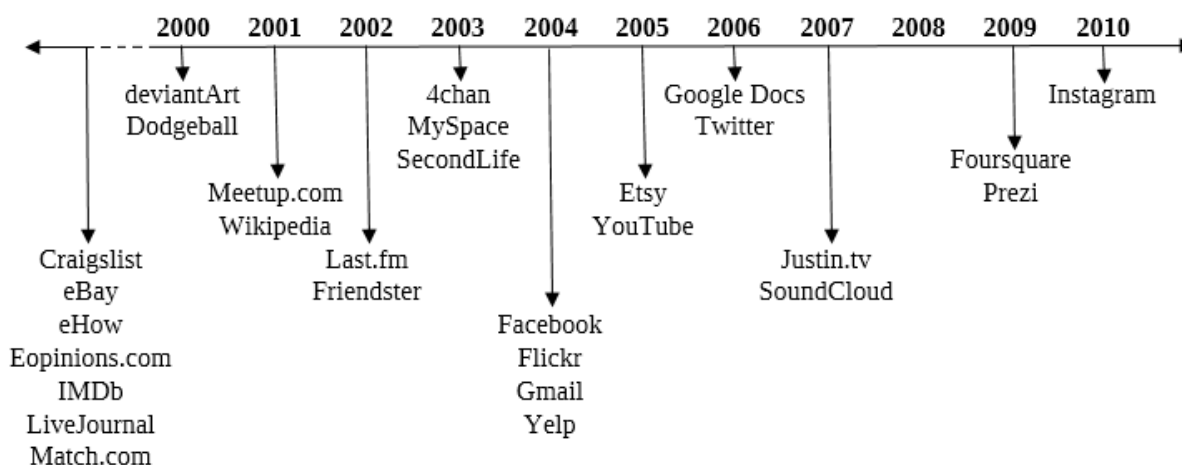
Nadaljnji razvoj tehnologij bo še naprej spreminjal značilnosti naravne rabe jezika na internetu. Baron (2003: 23) je denimo predvidela, da se bo spremenilo razmerje med informacijami v pisni in grafični obliki, z razvojem tehnologij za prepoznavo govora pa bo mogoče govorjena besedila spreminjati v tekst. Takšen način vnosa bo lahko vodil do radikalno drugačnih vrst računalniško posredovanih sporočil; emotikoni in krajšave bi lahko izginili, črkovanje in postavljanje ločil oz. slovnična pravilnost pa bi se lahko zaradi uporabe programov celo izboljšali. Vse naštetu se je v zadnjem desetletju pravzaprav že v veliki meri uresničilo.

Kljub vsemu naštetemu pa nekateri avtorji opozarjajo pred tehnološkim determinizmom in podarjajo, da četudi tehnologija in »nove« komunikacijske tehnologije omogočajo določene možnosti za komuniciranje, te običajno niso tako zelo spektakularne ali revolucionarne (Thurlow in Mroczek 2011: xxiv). Čeprav ne zanikajo vpliva tehnologije na jezik, se raje kot na iskanje »tipičnih« značilnosti »spletne govorice« osredotočajo na kontekstualne in posamične analize, ki odkrivajo, kako se različni kontekstualni parametri oblikujejo in pojavljajo v posameznih oblikah računalniško posredovane komunikacije (Androutsopoulos 2006: 421).

2.2 Jezik na družbenih omrežjih (jezik uporabniško generiranih vsebin, splet 2.0)

Zanimivo in tudi z jezikoslovnega vidika pomembno področje, ki dobiva vse več pozornosti v novejših raziskavah jezika na internetu, so družbeni mediji, kot sta Facebook in Twitter. Ena od možnih definicij družbenih medijev je, da gre za internetne strani in storitve, ki spodbujajo družbene interakcije med udeleženci, kar med drugim vključuje družbena omrežja, forume, bloge, podcaste, virtualne svetove in strani, ki omogočajo deljenje vsebin, denimo YouTube za videoposnetke ali Flickr za fotografije (Page et al. 2014: 5). Pojem »družbeni mediji« je izjemno širok in zajema različne internetne strani in platforme, ki omogočajo ustvarjanje in vzdrževanje omrežij oziroma skupnosti, ki se oblikujejo z medsebojnim pošiljanjem sporočil in drugih vsebin, zato je mednje mogoče šteti tudi strani, kot sta Wikipedia ali TripAdvisor (Seargeant in Tagg 2014: 3).

S pojavom novih tehnologij, ki omogočajo deljenje informacij, se je spletno okolje tako spremenilo, da je zanj Darcy DiNucci leta 1999 kot prva uporabila naziv splet 2.0 (angl. *Web 2.0*). Herring (2012: 4) je opredelila splet 2.0 kot »spletne platforme, ki so postale priljubljene v prvem desetletju 21. stoletja ter ki združujejo uporabniško generirane vsebine in družbeno interakcijo, pogosto kot odziv na strukture in/ali (multimedijsko) vsebino, ki jo zagotavljajo te spletne strani«. Nekateri uporabljajo izraza »družbeni mediji« in »splet 2.0« kot sinonima, drugi pa ju ločujejo (npr. Page et al. 2014: 8), vendar natančno razlikovanje za domet naše raziskave ni bistveno. Ključno pa je, da splet 2.0 odraža premik proti uporabnikom spleta, ki postajajo ustvarjalci (niso več zgolj potrošniki), spletne objave pa se hitro in neprestano posodablajo. Na sliki 1 je prikazan razvoj izbranih spletnih strani, ki so zaznamovale splet 2.0.



Slika 1: Časovnica spleta 2.0 z izbranimi spletnimi stranmi (povzeto po Herring 2012: 3).

Komunikacija, ki poteka na družbenih omrežjih, se od drugih spletnih žanrov oziroma množičnih medijev razlikuje po dvosmerni povezavi med udeleženci, pri čemer je poudarjen njihov dialoški potencial in možnost sodelovanja. Družbeni mediji pošiljatelju in prejemniku namreč omogočajo neposredno medsebojno interakcijo, pogosto v kratkem časovnem okviru, uporabniki pa se za takšno komunikacijo odločajo iz najrazličnejših vzrokov (Page et al. 2014: 12). V primerjavi s prvim valom raziskav internetnega diskurza avtorji v novejših objavah poudarjajo, da je težko vztrajati pri binarnem razlikovanju med sinhronimi in asinhronimi oblikami komunikacije, saj razlike med enimi in drugimi niso več povsem jasne (ibid.: 19). Tipičen primer omrežja, kjer je ta meja zabrisana, je tudi Twitter, ki je podrobneje predstavljen v nadaljevanju v razdelku 2.2.1.

Razvoj družbenih medijev, ki omogočajo povezovanje ljudi, je ustvaril nove in zanimive semiotične pritiske na jezik (Zappavigna 2012: 2). Tovrstne spletne strani namreč omogočajo alternativne načine komunikacije, ki prinašajo drugačno dinamiko in strukture, zato imajo pomemben vpliv na jezikovne in komunikativne navade (Seargeant in Tagg 2014: 2). Na rabo jezika na družbenih omrežjih recimo vplivajo različni konteksti in pričakovana zasebnost. Študija primera, ki sta jo izvedla Tagg in Seargeant, je pokazala, kako večjezični govorci na Facebooku preklaplajo med jeziki, ko želijo doseči določeno občinstvo, pri čemer ustvarjajo in vzdržujejo različne skupnosti v okviru iste spletne strani (Page et al. 2014: 22–23). V besedilih računalniško posredovane komunikacije so prisotni tudi interaktivni elementi, sicer značilni za govorjeni diskurz, kot so deiktični izrazi, elementi modalnosti ter referencialni in ekspresivni izrazi, še posebej značilni za tvite (Zwitter Vitez in Fišer 2015: 90).

Med splošnimi značilnostmi, ki so po opazanjih raziskovalcev izrazito značilne za družbene medije, so nestandardna raba ločil in velikih začetnic, črkovanje, besedišče in skladnja (Eisenstein 2013: 359), vendar je zgolj relativno majhen delež jezikovnih napak posledica nepozornosti ali neznanja (Herring 2001: 617). Uporabniki družbenih medijev uporabljajo tudi kodno preklapljanje med različnimi jeziki ali njihovimi različicami oziroma preklaplajo med različnimi registri, tako da uporabijo različno formalen način pisanja (Page et al. 2014: 31).

Tudi zaradi tovrstnih sprememb je jezik na družbenih omrežjih v medijih in javnosti pogosto označen za grožnjo jeziku oz. angleščini (Thurlow 2017: 7), mešanje kodov in kodno preklapljanje pa naj bi ogrožalo manjšinske jezike (Lenihan 2011: 58). Kljub temu je bil vpliv interneta na značaj posameznih jezikov do sedaj zelo omejen, večinoma pa se posamezne besede, slovnične konstrukcije in ortografski vzorci, ki se pojavljajo v besedilih na internetu, ne razlikujejo bistveno od tistih, ki nastanejo zunaj elektronskih medijev (Crystal 2011: 57).

Vsekakor bodo tudi v prihodnje potrebne vedno nove raziskave, ki bodo spremljale razvoj novih medijev in jezikovnih navad, ki jih bodo prinesli s sabo.

2.2.1 Raziskave jezikovne rabe na Twitterju

Twitter je zagotovo eden izmed vplivnejših novih medijev v zadnjem desetletju, zato ne preseneča, da so predmet mnogih raziskav, ki se ukvarjajo z jezikom na družbenih omrežjih, prav besedila s te platforme. Eisenstein (2013: 356) med razlogi za pogosto obravnavo Twitterja izpostavlja dejstvo, da so objave na njem privzeto javne, zato jih je lažje zbrati in uporabiti, kar je prednost predvsem v primerjavi z elektronsko pošto in zasebnimi sporočili na drugih omrežjih. V primerjavi z javnimi komentarji, klepetalnicami in forumskimi objavami pa omogoča lažje zbiranje podatkov, saj po nekaterih ocenah zgolj 7 % računov, ustvarjenih na Twitterju, uporablja nastavitve zasebnosti (Page 2014: 53). Kljub temu da so določene vsebine javne, je treba pri raziskovanju upoštevati omejitve, ki jih določa Zakon o varstvu osebnih podatkov (ZVOP-1).

Twitter, ki je zaživel leta 2006, je platforma, ki omogoča pošiljanje in prejemanje besedilnih sporočil (tvitov, angl. *tweets*), dolgih do 140 znakov. Tviti so prikazani na avtorjevi profilni strani, avtomatsko pa jih prejmejo avtorjevi sledilci (angl. *followers*) (Crystal 2011: 36). Poleg tega Twitter omogoča pošiljanje neposrednih sporočil (angl. *direct messages*), ki so namenjena zasebnim pogovorom med uporabniki in še niso dobro raziskana.

Tvite je mogoče brati, pošiljati in prejemati na različne načine – prek spletnih vmesnikov, e-pošte, SMS-ov ali tretjih programov, pogosto na mobilnih napravah (Zappavigna 2012: 3). Sčasoma so na Twitter dodali nove funkcionalnosti, kot sta uporaba simbola @ za naslavljanje drugih uporabnikov oz. kot referenco nanje ter # za semantično označevanje (Crystal 2011: 38). Znak za lojtro, zapisan skupaj z besedo, je znan kot hešteg ali ključnik (angl. *hashtag*), s katerim avtor označi oz. povzame temo svojega besedila. Heštegi so pravzaprav oblika metapodatkov, ki so vključeni v samo sporočilo ter omogočajo lažje iskanje (Zappavigna 2012: 1).

Zaradi omejenega prostora in izjemno dinamičnega okolja, v katerem nove objave hitro ostanejo spregledane, so uporabniki pod pritiskom, da za prenašanje sporočil uporabijo obstoječa sredstva na nove načine ali pa razvijejo nove načine izražanja, kot je denimo hešteg. Poleg tega morajo izraziti pomene, ki bi jih sicer lahko prenesli parajezikovno, na drugačen način, na primer z uporabo ločil ali velikih tiskanih črk (Zappavigna 2014: 139). Nadalje lahko tviti vključujejo povezave na druge multimedijske vsebine, zaradi prostorske omejitve pogosto v skrajšani različici (angl. *tiny URL*) (Zappavigna 2012: 3).

Dodatna prednost za raziskovalce je dejstvo, da Twitter zbira določene metapodatke o vsakem tvitu, npr. čas, ko je bil ustvarjen, ID uporabnika, kateremu je bil tvit namenjen (kadar je tak podatek na voljo), in informacije o uporabnikovem računu, vključno s številom sledilcev in predhodno objavljenih tvitov (Zappavigna 2012: 3). Poleg tega imajo uporabniki možnost, da dovolijo zbiranje drugih podatkov (npr. geolokacije) in uredijo svoj uporabniški profil, na podlagi česar lahko raziskovalci pridobijo dragocene sociodemografske podatke. Številne raziskave so zato posvečene odkrivanju povezav med jezikom na družbenih omrežjih in metapodatki, in sicer želijo raziskovalci zgolj na podlagi jezika oz. uporabljenih značilnosti ugotoviti denimo spol tvorca ali geografsko območje (Eisenstein 2013: 366).

David Crystal (2011: 39) je izpostavil, da jezik vsake nove oblike internetnega komuniciranja potrebuje nekaj časa, preden se ustali, ter da glede na to, da je Twitter relativno mlad medij (v času nastajanja njegove knjige je obstajal 4 leta, v času pisanja tega dela pa dobrih 10 let), lahko rezultati raziskav vsebujejo veliko individualnih variacij. Tudi zato so ugotovitve kakršnekoli analize močno odvisne od zajetega vzorca, ki mora biti dovolj velik in predvsem dobro vzorčen, da lahko sklepamo o posplošenih stilističnih značilnostih. Rešitev v zvezi s tem ponuja korpusno jezikoslovje, ki temelji na velikih količinah zbranih besedil, in prav ta pristop smo uporabili tudi v okviru naše raziskave.

Pri pripravljanju vzorca za analizo je treba upoštevati tudi to, da zaradi retvitanja prihaja do ponavljanja podatkov, kar je samo po sebi pomembna slogovna značilnost Twitterja (ibid.: 40–41). Uporabniki pri ponovni objavi tvita, ki ga je napisal drug uporabnik, pogosto uporabijo angleško kratico RT (angl. *retweet*), ki nakazuje, da je del tvita, ki sledi, pravzaprav citirano besedilo (Zappavigna 2012: 35). Status tovrstnih objav znotraj posamezne raziskave je tako pomembno teoretično vprašanje, saj se mora raziskovalec odločiti, ali jih bo vključil ali ne – odvisno od pojava, ki ga želi proučiti (Page et al. 2014: 175).

Nenazadnje pa je treba upoštevati, da je Twitter zasebno podjetje, ki lahko kadar koli spremeni funkcionalnosti aplikacij in pravila uporabe tako za uporabnike kot za raziskovalce. Iz preprostih besedilnih sporočil so se tviti sčasoma tako razvili v objave, v katere je mogoče vključiti fotografije, videoposnetke, ankete in celo prenose v živo, poleg tega so lahko zaradi rahljanja določenih pravil glede dolžine (npr. tako da se @imena in priloge ne štejejo v omejitvev 140 znakov) tviti tudi vse daljši (Sherman 2016).

2.3 Jeziki v stiku

Pomemben vpliv na rabo jezika na internetu oziroma družbenih omrežjih ima dejstvo, da so na svetovnem spletu prisotni številni jeziki. Kot je zapisal David Crystal: »Internet nudi dom vsem jezikom – brž ko imajo skupnosti dostop do elektrike in delujoče računalniške tehnologije. Njegov čedalje bolj večjezikovni karakter je najopaznejša sprememba od začetkov interneta, ko je bil povsem angleški medij« (2011: 78, prevod avtorice). Raziskovalci so že leta 1995 ugotavljali, da jeziki na internetu prihajajo v stik na dotlej neprimerljivi ravni (Paolillo 2007: 424), zato je denimo Paolillo že leta 1999 poudaril nujno potrebo po preučevanju stikov med jeziki na internetu in izjemno priložnost, ki jo splet zaradi velike količine razpoložljivih podatkov ponuja za preizkušanje teorij in ugotovitev iz realnega (nespletnega oz. nepovezanega) sveta. Tudi novejša raziskave ponujajo bogate dokaze, da računalniško posredovano komunikacijo zaznamujeta večjezičnost in kodno preklapljanje (Androutsopoulos 2013a: 3).

Stik med jeziki je mogoče najpreprosteje opredeliti kot uporabo več kot enega jezika na istem mestu in v istem času (Thomason 2001: 1). Z drugimi besedami gre za situacije, kjer pride do komuniciranja med govorce različnih jezikov, vendar za to ni potrebno, da gre za dvo- ali večjezične osebe, prav tako ni nujno, da so govorce prisotni na istem kraju (ibid.: 2). Stik med jeziki pogosto vključuje neposredno interakcijo med govorce, od katerih vsaj nekateri govorijo več kot en jezik, toda z razvojem interneta so se oblikovale številne nove priložnosti tako za učenje tujih jezikov kot njihovo uporabo na daljavo.

Začetnika raziskav o jezikih v stiku sta bila Uriel Weinreich (1953) in Einar Haugen (1950), v naslednjih desetletjih pa so to področje preučevali številni akademiki, ki so opravili poglobljene študije primerov in dognali kompleksne analitične okvire (Muysken 2010: 265). Dejstvo je, da so se jeziki oziroma njihovi govorce srečevali že tisočletja, najverjetneje vse od začetka človeštva. Prvi pisni dokazi za to segajo v antični čas, primere pa je mogoče najti tudi iz vseh obdobj do današnjega dne (Thomason 2001: 6–8). Myers-Scotton (2002: 31–32) našteva šest dejavnikov, ki spodbudijo ljudi k učenju novega jezika: vojaški vpadi in kolonizacija, življenje na obmejnem območju ali v etno-jezikovni enklavi, selitve iz socialnih ali ekonomskih razlogov (ne nujno prostovoljne), izobraževanje, širjenje mednarodnih jezikov (npr. latinščina v srednjem veku ali angleščina kot sodobna lingua franca) ter narodnostna zavest. Tudi Sankoff (2001: 3) poudarja pomen družbenih dejavnikov za stike med jeziki, ki so zgodovinsko gledano pogosto potekali v razmerah družbene neenakosti, ki je bila posledica vojn, osvajanj, kolonializma, suženjstva in migracij, pa tudi v enakopravnejših okoliščinah, npr. zaradi urbanizacije ali trgovanja. Jezikovne stike omogoča tudi tehnologija, predvsem internet, ki je

spodbudil nadaljnje širjenje angleščine po vsem svetu, pri čemer večino uporabnikov informacijsko-komunikacijske tehnologije predstavljajo govorci, katerih materni jezik ni angleščina (Wei 2000a: 3).

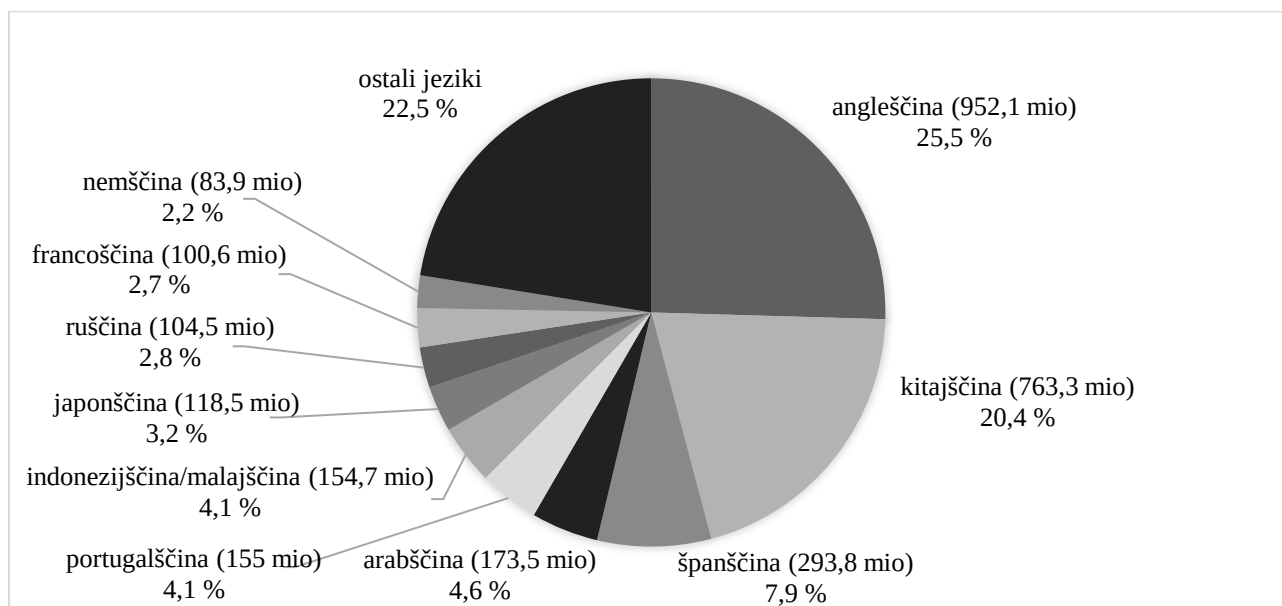
Zaradi vpliva drugih jezikov se jeziki spreminjajo na različne načine. Do kakšnih sprememb bo prišlo zaradi stika, je odvisno od družbenih dejavnikov, med drugim od intenzivnosti stikov (intenzivnejši kot so, več je možnih interferenc) ter števila govorcev posameznega jezika in njihovega družbenoekonomskega vpliva (jezik manjše skupine je bolj izpostavljen vnašanju značilnosti dominantnejšega jezika) (Thomason 2001: 66). Nekateri raziskovalci so ugotovili, da obstajajo določene omejitve pri prehajanju značilnosti iz enega jezika v drugega. Myers-Scotton je odkrila sistematične asimetričnosti med dominantnim matričnim (angl. *Matrix Language*) in vrinjenim jezikom (angl. *Embedded Language*) pri znotrajstavčnem kodnem preklapljanju, Poplack, Sankoff in Miller (1988) so ugotovili, da obstajajo precej pravilni statistični vzorci glede tega, katere leksikalne kategorije postanejo izposojenke (Muysken 2010: 267). Nasprotno Thomason trdi, da je vedno mogoče najti nasprotne primere, ki dokazujejo, da se lahko prenašajo najrazličnejše kategorije. Do določenih sprememb lahko pride že zaradi občasnih stikov, druge pa nastanejo le v okoliščinah, kjer je velika govorcev dvojezičnih. To pomeni, da si ni težko izposoditi nekaj besed iz jezika, ki ga govorec sploh ne govori dobro, za prevzemanje skladenjskih ali morfoloških struktur pa je treba precej obvladati izvirni jezik (ibid.: 78). Thomason posebej izpostavlja tudi vpliv učenja tujega jezika oziroma odsotnosti le-tega – v prvem primeru govorci prevzemajo manjši delež besedišča in več strukturnih značilnosti, v drugem pa ravno obratno (začnejo uporabljati več prevzetih besed) (ibid.: 128).

2.3.1 Vpliv angleščine na druge jezike

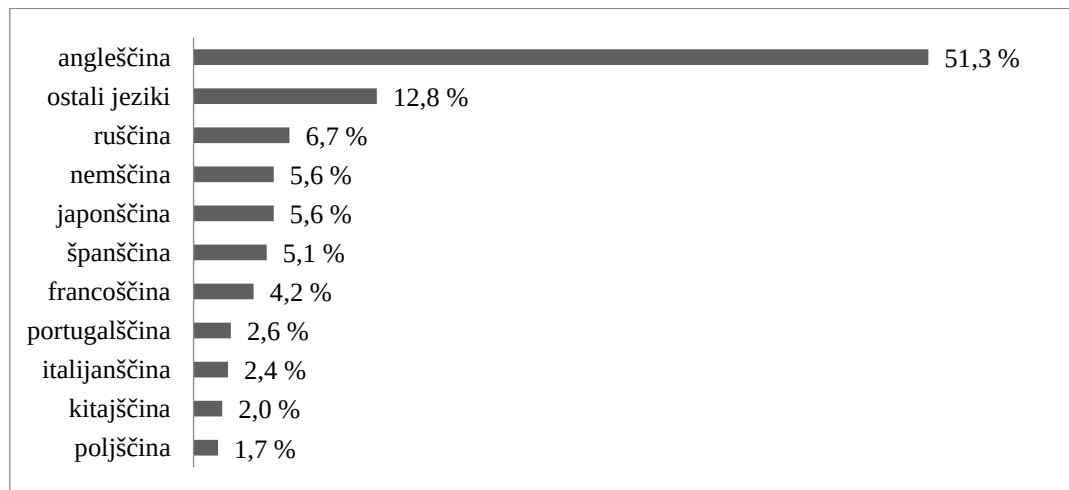
Številne študije o večjezikovnosti na internetu vključujejo dvojezičnost na ravni posameznika ali širše skupnosti (naroda) v kombinaciji z angleščino (Paolillo 2007: 425). Angleščina je bila prvi jezik na internetu in še danes prevladuje v številnih spletnih kontekstih, poleg tega po vsem svetu raste zanimanje za angleščino kot drugi jezik (Danet in Herring 2007: 22). Politični, gospodarski in ekonomski dejavniki so prispevali k temu, da je angleščina postala globalna lingua franca (Crystal 2003), zaradi njenega trenutnega statusa kot obveznega jezika na višjih stopnjah izobraževanja po vsem svetu pa bo angleščina tudi v bližnji prihodnosti zagotovo (ostala) v stiku z večino svetovnih jezikov (Thomason 2001: 24).

Vpliv angleščine na druge jezike se širi tudi s pomočjo spleta. Zaradi številnih tehnoloških omejitev sicer ni mogoče natančno ugotoviti zastopanosti jezikov na internetu (cf. Gerrand 2007), vendar kljub pomanjkljivostim različnih metod merjenja ni dvoma, da je angleški jezik

eden od prevladujočih, kar je razvidno tudi iz grafov 1 in 2. Kljub temu da angleščina nima več povsem dominantnega položaja, se do danes (še) niso uresničile nekatere napovedi, da bo kitajščina zamenjala angleščino kot prevladujoč jezik na internetu (Paolillo 2007: 421).



Graf 1: Prisotnost jezikov na spletu glede na število internetnih uporabnikov (Internet World Stats 2017).



Graf 2: Razmerje jezikov na spletu glede na količino vsebin (W3Techs 2017). Prikazanih je 10 najbolj uporabljenih jezikov, kot zanimivost pa lahko izpostavimo, da se je med 38 jezikov, za katere je bil zaznan vsaj 0,1-odstotni delež vsebin, uvrstila tudi slovenščina.

Med stereotipi, ki veljajo v povezavi s slovenščino na internetu, lahko zasledimo prepričanje, da so besedila slogovno slaba in slovnično nepravilna, pri čemer so opazni močni vplivi tujih jezikov (predvsem angleščine) in vključevanje prvin tujih jezikov, zaradi česar se govori o ogroženosti slovenščine (Michelizza 2007: 206–207). Pri tem se avtorji najpogosteje osredotočajo na prisotnost in vpliv angleščine, vendar ne gre zanemariti večjezičnosti interneta

ter da slovenski uporabniki pri pisanju posegajo tudi po elementih iz drugih jezikov, ki jih poznajo (vsaj do neke mere). Odstiranje dejanske rabe tujih jezikov je tudi eden od ciljev naše raziskave, vendar se glede na hipotezo, da bodo v korpusu tvitov prevladovali preklopi v angleščino, na tem mestu posvečamo samo temu jeziku, ki tudi glede na večino objav »trenutno v svoji vlogi globalnega jezika najmočneje zaznamuje podobo slovenščine in njen bodoči razvoj« (Šabec 2014a: 395).

Vpliv angleščine se je močno povečal zaradi ameriške popularne kulture, k njeni priljubljenosti pa so prispevali predvsem mediji (Šabec 2014a: 399). Ko gre za jezik računalniško posredovane komunikacije, je dominanten status angleščine razviden že iz prevzetih terminov, kot so blog, net, mail, link in drugi (Šabec 2009: 35). Slovenščino na spletu poleg tega zaznamuje odprtost za angleške vplive na vseh jezikovnih ravneh od besedja do skladnje in pravopisa (Šabec 2014a: 402). Dokaz za aktualnost te tematike so tudi številna diplomska in magistrska dela ter članki, ki se npr. ukvarjajo z vplivom angleščine na slovenščino v reklamnih oglasih (Štefančič 2016), na slovenščino na družbenih omrežjih (Laslo 2016), na sodobno slovensko poslovno terminologijo (Šabec 2014b), na spletno oglaševanje (Markelj 2015) in številna druga področja.

Nada Šabec (2009: 37–39) je analizirala mešanje slovenščine in angleščine na slovenskih blogih in med drugim ugotovila, da avtorji uporabljajo prilagojeno ortografijo (izpuščajo strešice ali uporabljajo angleške črke x, y, w), v določenih primerih spreminjajo besedni red pod vplivom angleščine (predvsem ko gre za rabo samostalniških zvez kot levih prilastkov) in uporabljajo tako znotrajstavčne kot medstavčne preklope v angleščino. Pri tem je posebej izpostavila dejstvo, da raba angleščine nima samo komunikacijske funkcije, temveč tudi družbeno, saj avtorji z njo kažejo, da so del določene spletne skupnosti.

Kot omenjeno zgoraj so raziskave o vplivu angleščine dokaj številčne, a zelo razdrobljene, zato zaenkrat ni mogoče podati točne ocene, kakšen je njen dejanski vpliv na slovenščino na splošno. Če si izposodimo besede Polone Gantar (Stabej et al. 2016: 103), da »ne moremo govoriti o 'slovarju spletne slovenščine', ampak da smo na veliko varnejših tleh, če govorimo o 'slovarju tviterščine', 'slovarju forumščine', se pravi besedil oz. žanrov, ki so vključeni v korpus Janes,« in jih nekoliko preoblikujemo, lahko zaključimo, da je praktično nemogoče opredeliti vpliv angleščine na (spletno) slovenščino, temveč lahko ta vpliv presojava zgolj v okviru posameznih besedilnih tipov oziroma kontekstov. Del k mozaiku dinamične slike sprememb v slovenskem jeziku pa bo, upamo, dodala tudi pričujoča raziskava.

2.4 Kodno preklapljanje

Kodno preklapljanje (angl. *code switching*) opisuje primere, ko dvojezični govorec uporabi več jezikov znotraj enega pogovora ali stavka (Gardner-Chloros 2009: 4), pri čemer vsaj enega od jezikov dobro obvlada, drugega pa bodisi zna uporabljati enako dobro kot prvega bodisi ga zna govoriti le deloma (Myers-Scotton 2002: 25). Pojavlja se v različnih vrstah situacij in je povezano tako z menjavo jezika (angl. *language shift*) kot z jezikovnimi spremembami (angl. *language change*) (Gardner-Chloros in Weston 2015: 188). V najširšem pomenu ljudje, tudi če so enojezični, pogosto preklaplajo med jezikovnimi različicami, da izrazijo več kot zgolj goli pomen besed, pri čemer lahko izbirajo med dialekti, registri, stopnjami formalnosti ali intonacijami (Gardner-Chloros 2009: 4).

Pri podajanju natančnejše opredelitve oziroma že pri samem poimenovanju kodnega preklapljanja sta nekoliko problematična oba dela termina, tj. »kod« in »preklapljanje«, poleg tega pa so različni avtorji za isti pojav uporabljali tudi druge izraze, denimo kodno mešanje (angl. *code mixing*) ali alternacijo koda (angl. *code alternation*). Kod je na splošno nevtralni krovni pojem za jezike, dialekte ali registre (Gardner-Chloros 2009: 11), torej lahko kodno preklapljanje poteka tudi med narečji istega jezika.

Poplack (1980: 583) je definirala kodno preklapljanje kot menjavanje dveh jezikov znotraj enega diskurza, stavka ali konstituenta (angl. *constituent*), Myers-Scotton (2000: 132) pa ga je opredelila kot uporabo dveh ali več jezikovnih različic v enem pogovoru, pri čemer je različica lahko karkoli od genetsko nesorodnih jezikov do dveh slogov istega jezika (Myers-Scotton in Ury 1977: 5). Ena od pogostejše citiranih definicij je tudi Gumperzova (1982), kjer je kodno preklapljanje opredeljeno kot »nizanje izrekov, ki pripadajo dvema različnima slovničnima sistemoma, znotraj istega govora« (Boztepe 2003: 4, prevod avtorice).

Za namene tega magistrskega dela bomo uporabljali nekoliko ožjo opredelitev, saj nas bo zanimalo preklapljanje med različnimi jeziki, ne pa tudi med dialekti ali registri. Ne glede na širšo ali ožjo definicijo pa lahko ugotovimo, da kodno preklapljanje zadeva praktično vsakogar, ki je v stiku z več kot enim jezikom (Gardner-Chloros 2009: 4), kar seveda vključuje večjezične situacije na spletu. S proučevanjem tega pojava raziskovalci dobijo vpogled v jezik z različnih vidikov, pomembna raziskovalna tema pa je tudi že iz preprostega razloga, da se s preklapljanjem srečuje večina ljudi na svetu (Gardner-Chloros 2009: 18).

2.4.1 Razmejitev od kodnega mešanja in prevzetih besed

Avtorji pogosto razmejujejo kodno preklapljanje od kodnega mešanja in izposojenk (angl. *borrowing* ali *loanword*), čeprav se tovrstne razprave velikokrat začnejo ali zaključijo z besedami Carol M. Eastman (1992: 10), da je vsak poskus razlikovanja med kodnim preklapljanjem, kodnim mešanjem in sposojanjem obsojen na neuspeh.

Na tem mestu velja izpostaviti, da v tuji literaturi za razliko od slovenske ne razlikujejo med izposojenkami in tujkami. Prve običajno pomenijo povsem prilagojene prevzete besede, druge pa niso popolnoma prilagojene jezikovnim zakonitostim slovenščine (Toporišič 1992: 299–300, 334). Novejša opredelitev, ki jo najdemo v teoriji, da so izposojenke »ljudske«, prevzete iz sosednjih tujih narečij, tujke pa »knjižne«, ki so jih vpeljali izobraženci (Snoj 2005: 115), pa je povsem neprimerna za opisovanje prevzemanja besedišča v multijezičnem spletnem in kulturnem okolju, kakršnemu smo izpostavljeni v sodobnem času. Toporišič (2004: 131) je kot skupno ime za besede, ki so prišle v slovenščino iz tujih jezikov in niso dediščina praslovanščine ali slovenščine, predlagal termin prevzeta beseda, ki ga bomo tudi mi v nadaljevanju uporabljali kot krovni izraz za tujke in izposojenke.

Nekateri avtorji kodno mešanje in kodno preklapljanje obravnavajo kot sinonima (npr. Grosjean 1982). Kachru (1983), Singh (1985), Sridhar in Sridhar (1980) so kodno preklapljanje uporabljali zgolj za znotrajstavčne preklope, kodno mešanje pa za medstavčne preklope (Boztepe 2003: 4), za razliko od Muysken (2006: 149), ki uporablja termin kodno mešanje za vse primere kodnega preklapljanja znotraj stavka. Natančnejše razlikovanje teh dveh pojmov za našo raziskavo ni pomembno, zato se mu ne bomo posvečali podrobneje, temveč bomo v nadaljevanju uporabljali izraz kodno preklapljanje za vse primere prekopov, tako znotraj stavkov kot med posameznimi stavki.

Ločevanje med kodnim preklapljanjem in prevzetimi besedami je potrebno, da lahko pred začetkom empiričnega dela raziskave natančneje določimo, kaj bomo sploh opazovali – ali bomo razlikovali kodne preklope od prevzetih besed in če, kako. Nekatere zgodnejše študije so denimo iz kodnega preklapljanja povsem izključile enobesedne elemente, v drugih so kot kodni preklopi šteli zgolj posamezni *neprilagojeni* leksemi, integrirane besede pa so bile opredeljene kot izposojenke (Myers-Scotton 1997: 20–21).

Na splošno velja, da kot tujejezični elementi najpogosteje nastopajo posamične besede, običajno občnoimenski samostalniki (Myers-Scotton 1997: 180), vendar za zdaj ni konsenza oziroma splošno sprejete metode za razlikovanje med prevzetimi besedami in preklopi.

Eden od predlaganih kriterijev za ločevanje med prevzetimi besedami in kodnimi preklopi je morfofonemska integracija besed z jezikom, s katerim so obdane (Gardner-Chloros in Weston 2015: 196). Izposojenke prevzamejo morfosintaktične značilnosti jezika, v katerega so prevzete, pri čemer se prilagodijo morfološkim zahtevam in pravilnemu vrstnemu redu besed, od izvirnega jezika pa obdržijo zgolj etimologijo ali določene elemente fonologije. Nasprotno pa kodni preklopi ohranijo slovnične lastnosti izvirnega jezika (Poplack 2015: 919). Drug kriterij je, da so prevzete besede razširjene v celotni skupnosti in so tako na voljo tudi enojezičnim govorcem (ibid.: 920). Vendar se je v številnih jezikovnih kombinacijah izkazalo, da lahko tudi ad hoc »preklopljeni« samostalniki in glagoli enako kot uveljavljene izposojenke pridobijo morfofonemske lastnosti jezika, v katerem se pojavijo. Tretji kriterij za ločevanje upošteva sinonimnost besed v enem in drugem jeziku, in sicer naj bi prevzete besede pogosteje zapolnjevale leksikalne vrzeli v jeziku, v katerega so prevzete, kodni preklopi pa so zgolj dodatna možnost za izražanje poleg že obstoječega ekvivalenta, toda ta opredelitev ne upošteva številnih funkcij, ki jih lahko opravlja kodno preklapljanje (Gardner-Chloros in Weston 2015: 196).

Z drugačnega zornega kota kodno preklapljanje in izposojanje nista strogo ločeni kategoriji, temveč del kontinuuma možnih sprememb, ki nastanejo zaradi jezikovnega stika, pri čemer se kodni preklopi s pogosto rabo sčasoma integrirajo ter postanejo izposojenke (npr. Myers-Scotton 2002; Thomason 2001). Po mnenju nekaterih se vsaka prevzeta beseda najprej pojavi kot kodni prekop, sčasoma pa se nekateri od prekopov generalizirajo oz. širše uveljavijo med govorce jezika (Gardner-Chloros 2010: 195). Myers-Scotton (1997: 176) je kot merilo za ločevanje med preklopi in prevzetimi besedami predlagala absolutno in relativno pogostost, saj kriterij fonološke, morfološke in sintaktične integracije ni dovolj natančen (ibid.: 191), pri čemer pa že sama izpostavlja dve težavi. Prva je absolutno število pojavitev (kako ločiti izposojenko od preklopa, če se obe obliki pojavita le enkrat ali dvakrat), druga pa arbitrarnost pri določanju, kako visoka relativna frekvenca je relevantna kot merilo (ibid.: 204).

Vendar je tudi glede prehajanja iz prekopov v izposojenke mogoče najti povsem nasprotna mnenja, tj. da je jezikovna integracija takojšnja, ne postopna, in da kodni preklopi ne postajajo prevzete besede, temveč govorce odločitev o tem, kateri pristop za vključitev tujejezičnega elementa bodo izbrali, sprejmejo v hipu (Poplack in Dion 2012: 279). Zaključimo lahko, da je za obe hipotezi mogoče najti empirične dokaze v določenih jezikovnih kombinacijah in okoliščinah, saj je različnih situacij (glede na jezike v stiku, govorce in njihovo znanje, osebne preference, družbene dejavnike itd.) praktično nešteto. Zaradi tega se glede na obseg pričujoče

naloge ne bomo podrobneje ukvarjali s teoretični merili za ločevanje prevzetih besed od preklopov, temveč smo se po vzoru podobnih raziskav (npr. Rosenberg 2013 ali Deuchar 2006) odločili, da bomo uveljavljenost prevzetih besed določili glede na njihovo (ne)vklučenost v jezikovne priročnike. Razlikovanje med kodnimi preklopi, ki bodo predmet raziskave, in prevzetimi besedami, ki jih ne bomo opazovali, tako temelji na predpostavki, da so prevzete besede do te mere integrirane in pogosto uporabljane v jeziku, v katerega so prevzete, da so na voljo tudi enojezičnim govorcem.

2.4.2 Vrste kodnega preklapljanja

2.4.2.1 *Situacijsko in metaforično preklapljanje*

Pogosto citirano dihotomijo med situacijskim in metaforičnim kodnim preklapljanjem sta razvila Jan-Petter Blom in John J. Gumperz na podlagi opažanj v bidialektalni skupnosti na Norveškem (Wei 2000b: 55). Pri situacijskem preklapljanju gre za neposredno povezavo med jezikom in družbeno situacijo, ko govorci izberejo določeno jezikovno obliko glede na svoje dojemanje dogodka, sprejete družbene norme oziroma kot odziv na določeno spremembo v okoliščinah, npr. kadar se skupini govorcev pridruži neznana oseba (Blom in Gumperz 2000: 116). Metaforično preklapljanje pa opisuje situacije, kjer ne pride do nobene bistvene spremembe okoliščin, temveč želijo govorci na ta način doseči posebne komunikacijske učinke. Blom in Gumperz kot primer navajata uradnike, ki uporabljajo standardno različico za pogovarjanje o uradnih zadevah, kadar z znanci govorijo o družinskih zadevah, pa znotraj iste situacije preklopijo v dialekt (ibid.: 117). Za našo raziskavo to razlikovanje ni relevantno, saj kljub določenim podobnostim jezika na Twitterju z govorjenim jezikom ni mogoče povleči primernih vzporednic s pogovornimi situacijami, kot jih opisujeta Blom in Gumperz. Mogoče bi bilo sicer opazovati rabo jezika istega uporabnika v različnih družbenih omrežjih oz. kako se le-ta spreminja glede na naslovnik (t.i. analiza družbenih omrežij, angl. *social network analysis*), vendar ta pristop ni bil uporabljen v okviru korpusa Janes.

2.4.2.2 *Znotrajstavčno in medstavčno preklapljanje*

Za našo raziskavo je relevantnejša delitev kodnega preklapljanja na znotrajstavčno (angl. *intrasentential*) in medstavčno (angl. *intersentential*) (Myers-Scotton 1997: 3). Podobno razlikovanje so uporabili tudi Poplack (1980) ter Appel in Muysken (2005). Na tem mestu velja ponovno nameniti nekaj besed terminologiji, ki jo bomo uporabljali tudi v raziskovalnem delu. V slovenščini je uveljavljeno razlikovanje med stavki (»enota ob osebni glagolski obliki«) in povedmi (»samostojna enota besedila«), pri čemer je prvi termin ekvivalenten angleškemu

»clause«, drugi pa »sentence«, torej bi bilo za »intrasentential« morda ustrežnejše uporabljati izraz »znotrajpovedno«. Kljub temu iz pragmatičnih razlogov v nadaljevanju uporabljamo stavek v skladu s *Smernicami za označevanje korpusa slovenskih tvitov Janes* (2016) kot »najmanjšo samostojno enoto jezikovnega sporočila, ki bi se v standardni slovenščini začela z veliko začetnico in zaključila s končnim ločilom«.

Pri medstavčnem preklapljanju je cel stavek (ali več stavkov) tvorjen v enem jeziku, preden pride do preklopa v drug jezik, do znotrajstavčnega preklapljanja pa pride znotraj enega stavka ali njegovega konstituenta. Pri tem se lahko v enem stavku pojavijo tri vrste konstituentov, ki: a) vsebujejo material iz obeh jezikov znotraj enega konstituenta, b) so v celoti v enem jeziku oziroma c) so v celoti v drugem (Myers-Scotton 1997: 4). Če to ponazorimo s primeri iz raziskovalnega korpusa, je v primeru i. »fotoshopat« konstituent a), saj je sestavljen iz angleške osnove (photoshop) in slovenske končnice (-at). V naslednjem tvitu (primer ii.) je »perfect« v celoti angleški konstituent, ostali so v slovenščini.

[i.] Ne da se mi jih fotoshopat, ampak neki sem morala objaviti

[ii.] perfect, k jst sam spečem, pojem ne 😊.

Za znotrajstavčno preklapljanje je nadalje značilno, da sta v stiku slovnici dveh jezikov, odločitev govorcev, da bodo v diskurzu uporabili znotrajstavčni prekop, pa je odvisna od družbenih, psiholoških in strukturnih dejavnikov (Myers-Scotton in Jake 2000: 265–266).

Kot primerno enoto za preučevanje znotrajstavčnega preklapljanja oziroma omejitev, ki veljajo za mešane elemente, je Myers-Scotton predlagala vezniško zvezo (angl. *complement phrase*) (1997: 243), iz česar izhaja ugotovitev, da medstavčno preklapljanje vsebuje preklope med konstituenti, ki so enojezični (Myers-Scotton in Jake 2000: 266). Medstavčni prekop lahko tako poudari določeno poanto, zaznamuje spremembo v udeležencih pogovora, kaže na to, na koga je izjava naslovljena, ali pa gre za neposreden citat (Hughes et al. 2006: 12).

Nekateri avtorji poleg znotrajstavčnega in zunajstavčnega ločujejo še t.i. pristavčno preklapljanje (angl. *tag switching*), ki pomeni vstavljanje tujejezičnih elementov (pristavkov, pojasnjevalnih dodatkov, vzklikov) v enojezični diskurz na začetku ali koncu stavka, kar daje dvojezični karakter sicer enojezičnemu stavku, zaradi česar je Poplack (1980) to vrsto preklapljanja poimenovala simbolično (angl. *emblematic*) (Appel in Muysken 2005: 118). Za potrebe naše raziskave smo tovrstne preklope šteli v kategorijo znotrajstavčnega preklapljanja, saj smo pri določanju vrste preklopa upoštevali stavčne meje v skladu s *Smernicami za označevanje* (priloga A), za ločeno označevanje pristavčnih prekopov pa se nismo odločili.

2.4.3 Dosedanje raziskave

Zaradi dejstva, da mora govorec, ki želi preklapljati med jeziki, vsaj do neke mere znati najmanj dva jezika, so kodno preklapljanje najprej preučevali v okviru dvojezičnosti, bodisi pri posameznih govornikih bodisi v celotnih skupnostih. Kljub temu je kodno preklapljanje ostalo bolj ali manj neopaženo v raziskavah o bilingvizmu vse do dela Johna J. Gumperza in njegovih sodelavcev v 60. in 70. letih 20. stoletja (Gardner-Chloros 2009: 9). Raziskave, posvečene kodnemu preklapljanju, je mogoče v grobem razdeliti v tri skupine (ibid.: 10):

- sociolingvistične/etnografske,
- pragmatične/analiza pogovora,
- slovnične.

Že s hitrim pregledom literature ugotovimo, da obstaja izjemno pester nabor študij kodnega preklapljanja, saj se razlikujejo glede na preučevane okoliščine (v dvojezičnih skupnostih, narodnostnih manjšinah, med priseljenci različnih starosti ipd.), ali preklapljanje poteka v govornem ali pisnem jeziku, ali uporabljajo mikro ali makro pristop (npr. analiza pogovora ali preklapljanje v celotni skupnosti) oziroma kvantitativni ali kvalitativni pristop, ali raziskava poteka v naravnih ali eksperimentalnih situacijah, poleg teh in drugih spremenljivk pa so mogoče tudi najrazličnejše jezikovne kombinacije.

Večina jezikoslovcev, ki so se ukvarjali s kodnim preklapljanjem, se je posvečala predvsem vprašanju, zakaj govorniki izberejo eno jezikovno različico namesto druge, da povedo isto stvar, oziroma z družbenimi funkcijami kodnega preklapljanja (Myers-Scotton 2002: 10). Nekatere študije se ukvarjajo izključno s slovnično strukturo preklapljanja, pri katerih družbeni dejavniki niso bistveni oziroma pogosto niso niti omenjeni (Myers-Scotton 2002: 10). Slovnične analize so se predvsem na začetku osredotočale na odkrivanje morebitnih univerzalnih omejitev oziroma načel, po katerih poteka preklapljanje. Glavni razlog za sklepanje, da kodno preklapljanje ni naključen pojav, je bilo opažanje, da so določene vrste kodnega preklapljanja pogostejše kot druge (Gardner-Chloros 2009: 95–96).

V etnografskih raziskavah je avtorje zanimal tudi odnos govorcev do preklapljanja. Med zanimivimi ugotovitvami je denimo ta, da so ljudje označili preklapljanje kot posledico lenobe in ga niso odobraval, četudi so ga sami uporabljali, oziroma da se niti niso v celoti zavedali, kako pogosto ga uporabljajo (Gardner-Chloros 2009: 15). Kljub temu pa se zdi, da je kodno preklapljanje postopno vse bolj družbeno sprejemljivo v številnih okoliščinah, predvsem med mlajšimi generacijami (ibid.: 82, 172). Tudi v kontekstu preklapljanja na Twitterju oziroma na

internetu na splošno bi bila zanimiva raziskava o odnosu uporabnikov do takšne rabe različnih tujih jezikov, še posebej če bi dobili odgovore ljudi, ki tudi sami (morda nevede) uporabljajo tujejezične elemente, a to žal presega obseg naše raziskave.

2.4.4 Kodno preklapljanje v računalniško posredovani komunikaciji

Raziskovalci so kodno preklapljanje sprva preučevali v govorjenih besedilih, najpogosteje v spontanem neformalnem govoru, pa tudi v institucionaliziranih kontekstih, kot so učilnice in pisarne (Sebba 2012: 97). Zanimanje za preklapljanje v pisnih besedilih se je razvijalo počasneje, predvsem v povezavi z določenimi obdobji, kot sta antika in srednji vek, sčasoma pa so se raziskave od preučevanja večjezikovnosti v literaturi razširile tudi na druga področja, npr. v medicinskih ali poslovnih besedilih. (Gardner-Chloros in Weston 2015: 182–184). Razvoj interneta je nato ustvaril ogromen prostor, relativno prost normativnih omejitev, kjer je mogoče večjezično sporazumevanje, zato ne preseneča, da so se najnovejše študije kodnega preklapljanja v pisnih besedilih osredotočile prav na računalniško posredovano komunikacijo (Sebba 2012: 98, 100).

Večina teoretičnih modelov kodnega preklapljanja temelji na verbalno ustvarjenih podatkih, zato še ni povsem jasno, do katere mere jih je mogoče uporabiti za pisna besedila (Garner-Chloros in Weston 2015: 183). Vendar so zaradi pomanjkanja teorij, ki bi bile razvite posebej za kodno preklapljanje v pisnih besedilih, raziskovalci večinoma črpali teoretično podlago iz obstoječih ugotovitev o preklapljanju v govoru. Predvsem pri klasifikaciji kodnega preklapljanja in ugotavljanju, kakšni so motivi za posamezne preklope, se je večina študij oprla na modele, ki so jih razvili Gumperz, Myers-Scotton in Auer, čeprav nobeden od njih ni bil v osnovi zasnovan za pisne kontekste (Sebba 2012: 99).

John J. Gumperz velja za pionirja sociolingvističnih študij kodnega preklapljanja. Razlikoval je med situacijskim in metaforičnim preklapljanjem ter razvil tipologijo diskurzivnih funkcij kodnega preklapljanja. Njegova spoznanja so uporabljena v skoraj vseh socio-lingvističnih študijah o pisnem preklapljanju. Auer (1984) je razvil model za analizo pogovora (angl. *Conversation Analysis*), ki ga je mogoče prilagoditi za analizo interaktivnih podatkov, ki spominjajo na pogovor, vendar ni uporaben za neinteraktivna pisna besedila (ibid.). Caroline Myers-Scotton je razvila model zaznamovanosti (angl. *Markedness Model*), ki ga je mogoče vsaj do neke mere uporabiti za pisne žanre, ki so podobni pogovoru (Sebba 2012: 99). Njen drugi pogosto citiran teoretični model je okvir matričnega jezika (angl. *Matrix Language Framework*), ki bo podrobneje obravnavan v razdelku 2.4.5.

Glede na to, da sta kodno preklapljanje in kodno mešanje običajna pojava pri ljudeh, ki znajo dva ali več jezikov, sta zlahka našla pot v računalniško posredovano komunikacijo, vendar je kodno preklapljanje manj raziskano kot drugi jezikovni procesi v tem mediju (Androutsopoulos 2013b: 667), čeprav so se posamezniki ukvarjali s to tematiko že od 90. let 20. stoletja (npr. Paolillo 1996). Podobno kot velja za preklapljanje v pisnih besedilih na splošno, zaenkrat še ni razvite splošno sprejete metodologije, ki bi upoštevala specifične značilnosti jezika na internetu. Raziskovalci tako črpajo iz različnih okvirjev, razvitih za analizo govorjenega diskurza, pri čemer prevladujeta pragmatični in sociolingvistični pristop, ne toliko slovnični in lingvistični (Androutsopoulos 2013b: 668). Raziskave zajemajo raznolik nabor žanrov in sociolingvističnih okoliščin ter uporabljajo različne metode, tako kvalitativne kot kvantitativne (ibid.: 673), vendar je bil v večini raziskav uporabljen (zgolj) kvalitativni pristop, saj so se osredotočale na izražanje družbene identitete, ne pa na preverjanje hipotez na velikih korpusih (ibid: 679).

Raziskave kodnega preklapljanja v računalniško posredovani komunikaciji je glede na družbene okoliščine mogoče razdeliti na dva glavna sklopa, in sicer tiste, ki se ukvarjajo z izseljenci, diasporami in etničnimi manjšinami, ter tiste, ki se ukvarjajo s preklapljanjem med mlajšimi udeleženci in globalno-lokalno (»glokalno«) rabo jezika, ki je povezana z glasbo in medijsko kulturo. Tu gre predvsem za stik med angleščino in nacionalnim jezikom, pri katerem avtorji opazajo »minimalni bilingvizem«, tj. da govorci uporabljajo ustaljene angleške besedne zveze (npr. pozdrave, medmete in diskurzne povezovalce, slogane) (ibid: 678).

Androutsopoulos (ibid.: 674) je zbral najpomembnejše raziskave, v katerih so preučevali kodno preklapljanje na IRC-u, v elektronski pošti, na spletnih forumih in v SMS-ih, in sicer v javni in zasebni komunikaciji ter različnih jezikovnih kombinacijah, najpogosteje v paru z angleščino. Z razvojem novih medijev in družbenih omrežij se avtorji vse pogosteje odločajo tudi za preučevanje kodnega preklapljanja na Twitterju, Facebooku, blogih, v komentarjih in drugih spletnih okoljih. Veliko je tudi magistrskih in diplomskih nalog, ki se ukvarjajo z mešanjem jezikov v pisnih besedilih, zato lahko sklepamo, da je preučevanje kodnega preklapljanja v omenjenih novih medijih še kako aktualno. Študentje so denimo opazovali uporabo kodno preklapljanje v angleščino na Twitterju med maternimi govorci turščine (Tastan 2012), med hrvaškimi uporabniki Facebooka (Lončarić 2014), na švedskih forumih (Rosenberg 2014) in v finskih spletnih igričarskih videoposnetkih (Myllärinen 2014), med slovenskimi pa lahko izpostavimo preklapljanje v komentarjih na spletne članke (Krošelj 2015) in vpliv angleščine na slovenščino na družbenih omrežjih (Laslo 2016).

2.4.5 Okvir matričnega jezika (*Matrix Language Framework*)

Okvir matričnega jezika (angl. *Matrix Language Framework* oz. MLF), ki ga je razvila Carol Myers-Scotton, je najbolj celovit sistem, ki združuje ureditev slovnice pri dvojezičnih govorcih, slovnične omejitve kodnega preklapljanja in družbene dejavnike (Savić 1996: 43). Model temelji na predpostavki, da jezika, ki sta v stiku, nista enakopravna, temveč ima eden (t.i. matrični oz. glavni) dominantno vlogo in zagotavlja morfosintaktični okvir, v katerega je mogoče vstavljati elemente iz drugega jezika (t.i. vrinjeni jezik oz. jezik, ki prispeva tuje prvine), bodisi kot posamične elemente ali daljše zveze, ki jih Myers-Scotton imenuje »otoki vrinjenega jezika« (angl. *Embedded Language Islands*). Izpostaviti velja, da za razliko od naše raziskave Myers-Scotton v ta model ni vključila medstavčnih preklapov.

Joshi je že leta 1982 pokazal, da kodno preklapljanje zaznamuje asimetrija glede udeležbe vpletenih jezikov, kar je utemeljil z opažanjem, da se govorci in poslušalci običajno strinjajo o tem, v katerem jeziku je »mešan« stavek (Joshi 1982: 145). Myers-Scotton (1997: 35) je za identifikacijo matričnega jezika najprej predlagala kriterij pogostosti: »Matrični jezik je tisti, iz katerega prihaja več morfemov v vrstah interakcije, ki vsebujejo znotrajstavčno kodno preklapljanje«.

Na podlagi empiričnih podatkov se je pozneje izkazalo, da ta opredelitev ne drži vedno (ibid.: 246), zato sta Myers-Scotton in Jake (2000: 266–267) opisali dva strukturna kriterija za določanje matričnega jezika, in sicer sta to a) jezik, ki zagotavlja morfosintaktični okvir za stavek, ki vsebuje znotrajstavčni preklop, in b) jezik, iz katerega običajno prihaja večina morfemov v relevantnem vzorcu. Tretji kriterij je sociolingvistični – matrični jezik je nezaznamovana oziroma pričakovana izbira za komuniciranje v situaciji, v kateri se pojavi znotrajstavčno preklapljanje. Takšen nezaznamovan status je načeloma empirično dokazljiv, saj matrični jezik prispeva kvantitativno več jezikovnega materiala v celotnem diskurzu.

Stavki, v katerih pride do kodnega preklapljanja, lahko tako vsebujejo tri vrste konstituentov:

1. otoke matričnega jezika (MJ), ki vsebujejo samo morfeme iz matričnega jezika,
2. otoke vrinjenega jezika (VJ),
3. elemente MJ + VJ, ki vsebujejo material iz obeh jezikov znotraj enega elementa.

Prvo in tretjo skupino ureja slovnica matričnega jezika, za otoke VJ pa je značilno, da so tvorjeni v skladu s slovničnimi pravili vrinjenega jezika (Myers-Scotton 1997: 6). Za kodno preklapljanje morejo govorci poznati vsaj nekaj morfemov iz vrinjenega jezika, sicer pa ni potrebno, da obvladajo tuj jezik (ibid.: 8). Vsak od jezikov ima tudi drugačen sociolingvistični

status – vsaj v vrsti interakcije, kjer pride do kodnega preklopa, pogosto pa tudi v družbi na splošno (ibid.: 20)

Druga pomembna predpostavka modela je razlikovanje med sistemskimi in vsebinskimi morfemi. Termin »vsebinski morfemi« bolj ali manj ustreza polnopomenskim/leksikalnim besedam, »sistemski morfemi« pa funkcijskim/slovničnim besedam (ibid.: 48). Slednji namreč prispevajo semantična in pragmatična sporočila, ki jih želi prenesti govorec, sistemski morfemi pa pri tem delujejo kot slovnični sistem, ki prispeva obliko (ibid.: 255). Myers-Scotton sicer opozarja, da se te kategorije ne pokrivajo popolnoma, zaradi česar je vzpostavila svoj sistem za razlikovanje med sistemskimi in vsebinskimi morfemi, po katerih je lahko npr. »predlog« enkrat v prvi, drugič pa v drugi kategoriji (ibid.: 101–122), toda za domet pričujoče raziskave podrobnejše razlikovanje ni potrebno. Tvitov namreč nismo označevali glede na njeno klasifikacijo, temveč v skladu s Smernicami za označevanje (priloga A), kjer je predvidenih 11 besednih vrst in kategorija »neuvrščeno«.

Glavni cilj modela MLF je pojasniti, na kakšen način se lahko jezikovni material iz vrinjenega jezika pojavi v matričnem jeziku pri tvorjenju kodnih preklapov – ali se vsebinski morfem iz VJ pojavi kot posamični leksem (ki je lahko morfološko integriran v okvir matričnega jezika ali pa je v »goli« obliki) ali v otoku VJ oziroma ali se sploh lahko pojavi. Odgovore na to naj bi dale slovnične informacije na abstraktni ravni jezikovne kompetence, ne zunanje okoliščine, v katerih pride do preklapljanja (ibid.: 242).

2.4.6 Slovnične omejitve pri kodnem preklapljanju

Mnenje, da kodno preklapljanje urejajo splošna načela, ne partikularne omejitve, je sprejeto med mnogimi raziskovalci, toda med njimi ni soglasja, kakšna so ta pravila. Nekateri vztrajajo, da je iskanje tovrstnih splošnih načel brezplodno (npr. Gardner-Chloros), drugi trdijo, da omejitve sploh ne obstajajo, tretji pa menijo, da se morebitne omejitve razlikujejo med različnimi skupnostmi (npr. Muysken) (Poplack 2015: 919–920). V nadaljevanju naštevamo le nekaj najbolj razširjenih hipotez, na katere se bomo lahko oprli tudi pri naši raziskavi.

2.4.6.1 *Omejitev glede strukturne integritete (equivalence constraint, structural-integrity constraint)*

V 70. letih 20. stoletja so raziskovalci predvsem na podlagi preučevanja kodnega preklapljanja med španščino in angleščino prišli do splošne ugotovitve, da kodni preklopi niso mogoči v kontekstih, kjer prihaja do navzkrižja/konflikta v besednem redu med matričnim in vrinjenim jezikom (Myers-Scotton 1997: 26). Sankoff in Poplack (1981) sta predpostavila, da mora biti

skladnja na obeh straneh kodnega preklopa slovnično pravilna v zadevnih jezikih (Clyne 2009: 747). Toda v poznejših študijah so našli veliko nasprotnih primerov v različnih jezikovnih kombinacijah, s katerimi so ovrgli univerzalnost te omejitve (Myers-Scotton 1997: 28–30).

2.4.6.2 Omejitev glede prostih morfemov (*free-morpheme constraint*)

Sankoff in Poplack (1981) sta predvidela, da ne more biti do kodnega preklopa med vezanim morfemom in leksikalno obliko, razen če je leksikalna oblika fonološko integrirana v jezik, iz katerega prihaja vezani morfem (Clyne 2009: 747). Tudi za to omejitev so raziskovalci našli številke nasprotne primere, pri čemer Myers-Scotton (1997: 31–32) ugotavlja, da večina le-teh prihaja iz aglutinacijskih jezikov.

2.4.6.3 Omejitev glede možnosti preklapljanja zaprtih besednih vrst (*constraint on switchability of closed-class groups*)

V sistemu, ki ga je oblikoval Joshi, večina omejitev temelji na splošni omejitvi glede možnosti preklapljanja slovničnih oziroma zaprtih besednih vrst (določila, kvantifikatorji, predlogi, morfemi, ki izražajo glagolski čas, vezniki, zaimki, morfemi za izražanje svojine, pomožni glagoli) (Clyne 2009: 748).

Podobno načelo z nekoliko drugačno terminologijo (t.i. *System Morpheme Principle* oz. načelo sistemskih morfemov) je predlagala Myers-Scotton, in sicer da morajo sistemski morfemi (določila, števniki, svojilni pridevniki, kopule, vezniki, končnice za spol in sklon ter glagolski čas in vid), ki se pojavijo v elementih MJ + VJ, prihajati iz matričnega jezika (Myers-Scotton 1997: 120–121). To hipotezo je še zamejila s t.i. *Blocking Hypothesis*, ki pravi, da matrični jezik onemogoči pojavitev vsebinskih morfemov iz vrinjenega jezika, ki niso v skladu z določenimi pogoji kongruence oziroma se ne ujemajo z lemo iz matričnega jezika (Myers-Scotton 1997: 230).

V preučevanem korpusu iz Nairobija, ki je vseboval kodne preklope v svahiliju, so si denimo med angleškimi leksemi, ki so se pojavili posamično, besedne vrste po pogostosti sledile v naslednjem zaporedju (od največ do najmanj pojavitev): samostalniki, glagoli, pridevniki, prislovi, medmeti, vezniki, zaimki (ibid.: 15).

2.4.6.4 Hierarhija otokov vrinjenega jezika

Bolj kot je posamezen konstituent periferen znotraj stavka, bolj prosto se lahko pojavi kot otok vrinjenega jezika (Gardner-Chloros 2009: 45). Myers-Scotton je na podlagi korpusa iz Nairobija, oblikovala naslednjo lestvico, pri kateri so na vrhu tisti elementi, ki so najbolj

dostopni govorcem oziroma najlažje postanejo znotrajstavčni preklopi kot otoki vrinjenega jezika (1997: 144):

1. šablonski izrazi in idiomi (predložne zveze časa in načina; dopolnila glagolskih zvez)
2. drugi načini za izražanje časa in načina (prilastki samostalniških/predložnih zvez, uporabljeni kot prislovi)
3. kvantifikatorji (pridevniške in samostalniške besedne zveze, predvsem kot dopolnila glagolskih besednih zvez)
4. samostalniške besedne zveze in dopolnila glagolskih zvez, ki ne izražajo kvantifikacije ali časa (samostalniške zveze, pridevniške zveze, dopolnila)
5. samostalniške zveze, ki izražajo agenta,
6. morfemi, ki določajo tematske vloge in sklone, tj. glavne osebne glagolske oblike (spregane v celoti).

Za konec velja poudariti, da so novejša raziskava in predvsem raziskave preklapljanja med drugimi jezikovnimi kombinacijami takšne predpostavke ovrgle oziroma za vsako domnevno omejitev našle nasprotno primere, predvsem pa se je izkazalo, da hipoteze niso splošne, temveč kvečjemu veljajo le v določenih jezikovnih parih (Myers-Scotton 1997: 34). Thomason (2001: 151) v zvezi s tem poudarja, da so v primerih, ko prevzemanje določenih elementov iz enega v drug jezik glede na predpostavke ni mogoče, odločilni dejavniki na strani posameznika, ki lahko vselej poveže morebitne slovnične in druge omejitve. Zaradi tega je treba imeti v mislih, da vse predlagane omejitve niso kategorične, temveč zgolj odražajo tendence v določenih okoliščinah (Gardner-Chloros 2009: 154).

2.4.7 Funkcije kodnega preklapljanja

Kot že omenjeno, so bile klasifikacije funkcij kodnega preklapljanja razvite v okviru raziskav, ki so preučevale preklapljanje v govorjenih (neformalnih) besedilih, vseeno pa so uporabne tudi za opisovanje kodnih prekopov v računalniško posredovani komunikaciji. V nadaljevanju tako na kratko predstavljamo nekaj klasifikacij, ki predstavljajo podlago za kvalitativni del naše raziskave, v kateri bomo poskušali podati nekaj razlogov, zakaj ljudje preklaplajo na Twitterju oziroma v kakšnih okoliščinah posegajo po tujejezičnih prvinah.

Gumperz je denimo opredelil naslednje diskurzne funkcije kodnega preklapljanja: preklapljanje v odvisnem govoru, določanje naslovnika, pojasnjevanje, poudarjanje, izražanje ekspresivnosti, označevanje sporočil (ločevanje dejstev od komentarjev) ter ločevanje osebnih in objektivnih stališč (Androutsopoulos 2013b: 669).

Appel in Muysken (2005: 119–120) se naslanjata na šest funkcij govora, kot jih je opredelil Jakobson (referencialna, emotivna, poetska, fatična, metajezikovna in konativna funkcija), ter pravita, da so primerne tudi za preučevanje kodnega preklapljanja.

Gardner-Chloros (2009: 69) je različne pogovorne funkcije kodnega preklapljanja združila v naslednje kategorije: citati, določanje naslovnikov, medmeti, ponovitve, označevanje sporočil, personalizacija/objektivizacija.

Androutsopoulos (2013b: 681) je diskurzne funkcije kodnega preklapljanja v računalniško posredovani komunikaciji opredelil na podlagi klasifikacij, ki pojasnjujejo preklapljanje v pogovorih, in sicer naj bi ljudje preklapljali iz naslednjih razlogov:

- v formulaične diskurzivne namene, vključno s pozdravi, poslovnimi pozdravi in dobrimi željami,
- za tvorjenje kulturno specifičnih žanrov, kot so poezija ali šale,
- za navajanje odvisnega oz. poročanega govora (ki je v nasprotju z avtorjevo izjavo),
- za ponovitev izjave z namenom poudarjanja,
- za naslavljanje določenega naslovnika, kot odziv na jezikovno izbiro predhodnika ali z namenom upreti se jezikovni izbiri drugih udeležencev,
- da bi kontekstualizirali spremembo teme ali perspektive, podarili razliko med dejstvi in mnenjem, informativno in čustveno izjavo ipd.,
- z namenom zaznamovati, kaj je izrečeno v šali in kaj zares,
- s preklopi v jezik sogovornika ali iz njega pa lahko tudi izrazijo svoje strinjanje ali nestrinjanje.

V kvalitativnem delu raziskave se bomo oprli predvsem na klasifikaciji zadnjih dveh avtorjev, in sicer bomo poskusili ugotoviti, ali je mogoče tovrstne funkcije preklapljanja zaznati tudi v slovenskih tvitih. Na tem mestu lahko izpostavimo še opažanje, da na kodno preklapljanje vplivajo trije sklopi dejavnikov (Gardner-Chloros 2009: 42):

1. dejavniki, neodvisni od posameznega govorca in situacije (zadevajo namreč vse osebe v določeni skupnosti; npr. prestižnost jezika ali razmerja moči),
2. dejavniki, vezani na govorca (znanje jezika, socialna omrežja, ideologije),
3. dejavniki znotraj pogovora.

Verjetno je nemogoče ugotoviti, kolikšen je vpliv posameznega dejavnika v konkretni situaciji, saj je vsakršna uporaba jezika odvisna od številnih okoliščin in udeležencev. Učinek različnih družbenih spremenljivk na kodno preklapljanje na spletu še ni bil sistematično raziskan

(Androutsopoulos 2013b: 679), pogosto pa raziskovalce zanima odnos med kodnim preklapljanjem in ustvarjanjem identitete (npr. Seargeant in Tagg 2014).

Tudi v okviru tega magistrskega dela ne bo mogoče ugotoviti in primerjati vpliva posameznih dejavnikov, saj vse informacije, ki bi bile potrebne za sklepanje zaključkov, niso na voljo. Kljub temu pričakujemo, da bomo na podlagi predstavljenih primerov lahko ponazorili pester nabor okoliščin, v katerih slovenski uporabniki Twitterja posežejo po tujejezičnih besedah.

Myers-Scotton denimo trdi, da je motivacija za kodno preklapljanje predvsem družbena – govorci preklopijo v drug jezik, saj jim dodatni kod omogoča projiciranje druge dimenzije svoje osebnosti ali pa lahko na ta način vplivajo na odnose z drugimi udeleženci v pogovoru. Besedne zveze iz vrinjenega jezika lahko izberejo tudi zato, ker bolje izpolnjujejo govorčeve semantične/pragmatične namene, npr. ker se spremeni poudarek zaradi drugačnega besednega reda ali ko gre za preklope idiomatičnih fraz, katerih pomena ni mogoče povsem natančno prenesti v drug jezik. Kodno preklapljanje brez očitne motivacije pa je del posameznikove identitete (Myers-Scotton 1997: 248–251).

V slovenskem prostoru sta Mikolič Južnič in Čibej (2016: 28) vzroke za preklapljanje v elektronskih medijih povzela v naslednjih kategorijah:

- izkazovanje solidarnosti oz. pripadnosti skupini, izražanje identitete,
- naslavljanje izbranega občinstva,
- pomanjkanje izraza,
- pomanjkanje terminologije,
- ustaljene fraze,
- pisanje o določeni temi (npr. tabuji),
- izražanje čustev, odnosa (npr. neodobravanje, jeza, veselje),
- retorično sredstvo (prepričevanje bralcev, poudarjanje).

2.5 Korpus Janes

Korpus Janes je nastal v okviru nacionalnega raziskovalnega projekta JANES – Jezikoslovna analiza nestandardne slovenščine (<http://nl.ijs.si/janes/>) v sodelovanju raziskovalcev s Filozofske Fakultete Univerze v Ljubljani in Instituta Jožef Stefan. Gre za obsežen korpus spletnih uporabniških vsebin, ki vsebuje tvite, spletne forume, novice in uporabniške komentarje nanje, bloge in komentarje nanje ter uporabniške in pogovorne strani na Wikipediji. Poleg tega so besedila v njem jezikovno označena in opremljena z naborom metapodatkov (Fišer et al. 2016: 68–69)

V času nastajanja tega magistrskega dela je bila sicer objavljena nova različica – Janes 1.0, vendar je bil osnova za našo raziskavo korpus Janes 0.4, zato v nadaljevanju navajamo podatke zanj. V Janes 0.4 je vključenih več kot 9 milijonov besedil s skupno dobrih 200 milijonov pojavnic oz. 175 milijonov besed (ibid.: 73). Razdeljen je na 5 podkorpusev (tviti, forum, blog, news in wiki), od katerih je največji podkorpus tвитov z več kot 107 milijoni pojavnic oz. 90 milijonov besed (ibid.: 73).

Za korpus Janes 0.4 je značilno tudi to, da so besedila opremljena s številnimi metapodatki. Nekateri od njih so bili zajeti neposredno, predvsem URL-naslov, pri tvitih pa tudi uporabniško ime avtorja, datum in čas objave, število posredovanj (angl. *retweets*), všečkov (angl. *favourites*) in geolokacija, če je imel uporabnik vklopljeno tudi to funkcijo (ibid.: 75).

Za našo raziskavo je dragoceno tudi to, da je označen spol avtorja besedila, saj bomo preverili, ali morda prihaja do razlik pri kodnem preklapljanju med moškimi in ženskami, čeprav je naša hipoteza ta, da se raba med spoloma ne bo bistveno razlikovala. Spol avtorjev drugih besedil v korpusu je bil določen z uporabo oblikoskladenjsko označenega korpusa, za tvite pa je bil označen ročno (ibid.: 78–79). Drug relevanten metapodatek je standardnost besedil, saj pričakujemo, da bo več tujejezičnih prvin oz. kodnih preklapov v manj standardnih besedilih.

Označevanje (ne)standardnosti je potekalo z avtomatsko metodo (Ljubešić et al. 2015), za besedila pa sta ločeno označeni tehnična in jezikovna nestandardnost. Tehnična (T) je ožje pravopisna in se izraža predvsem v (ne)uporabi velikih začetnic, ločil, presledkov in prisotnosti tipkarskih napak, jezikovna (L) pa upošteva izbor in zapis besed, njihove oblikoslovne lastnosti in besedni red. Za obe je uporabljena trostopenjska lestvica (1 = povsem standardno, 3 = zelo nestandardno) (ibid.: 84).

2.5.1 Podkorporus tvitov

Zajem tvitov je celovit, kar pomeni, da so vanj vključeni vsi uporabniki in njihove objave, ki so bili identificirani z namenskim orodjem TweetCat, s katerim je mogoče izluščiti uporabnike, ki tvitajo v ciljnem jeziku (za korpus Janes je to slovenščina). Orodje uporablja Twitter Search API, da najprej prepozna uporabnike, ki potencialno tvitajo v ciljnem jeziku, nato pa izvede še natančnejšo identifikacijo jezika uporabnika, tako da mu ves čas sledi in sproti shranjuje objavljene tvite. Pred dokončnim zajemom podatkov v korpus Janes so raziskovalci izvedli še dodaten korak filtriranja uporabnikov, tako da so s Pythonovim modulom langid.py izločili uporabnike, katerih večinski jezik ni slovenščina. (ibid.: 70). Ker Twitter API vrača samo zadnjih 500 tvitov posameznega uporabnika, je pomembno, da se tviti zbirajo sproti, v Janes 0.4 pa so vključeni tviti od junija 2013 do januarja 2016 (ibid.: 76).

Ključno za našo raziskavo je prav to, da so bila vsa filtriranja narejena na uporabnikih, ne na tvitih, zato so v korpus vključeni tudi tviti, (delno) zapisani v drugih jezikih, če je njihov avtor preverjeno slovenski uporabnik. Z namenom, da bi omogočili izločitev tujejezičnih besedil, so bila vsa besedila v korpusu Janes avtomatsko označena s programom langid.py, ki prepozna jezik in poda oceno verjetnosti identificiranega jezika. Vendar so rezultati vsebovali precej šuma, zato so se avtorji korpusa odločili, da od vseh pripisanih oznak ohranijo le štiri: slv za slovenščino, eng za angleščino, hbs za hrvaščino, srbščino ali bosanščino oziroma und za nedoločen jezik (ibid.: 82).

Na kratko predstavljamo še nekaj podatkov o podkorporusu:

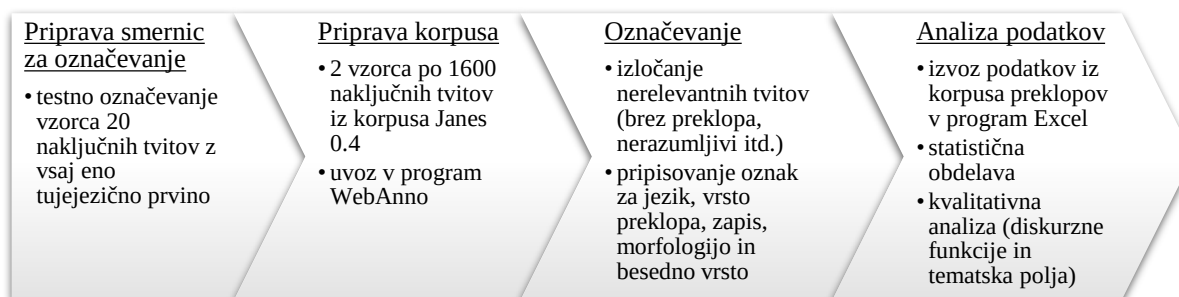
- število besedil: 7.503.199,
- število besed: 90.180.337,
- število pojavnic: 107.053.232,
- število uporabnikov oz. avtorjev: 8.749, od tega 53 % moških in 47 % žensk,
- 9,6 % tvitov je bilo identificiranih kot angleških.

3 Raziskovalni del

3.1 Načrt raziskave

Temeljni cilj magistrskega dela je raziskati značilnosti kodnega preklapljanja v spletni komunikaciji na Twitterju. S tem namenom smo na podlagi tvitov iz korpusa Janes 0.4 pripravili raziskovalni korpus tvitov, ki vsebujejo vsaj eno tujejezično prvino, in sicer v dveh korakih. Prvi vzorec je bil uravnotežen glede na spol in standardnost besedila (v nadaljevanju »splošni vzorec«), nato pa smo na podlagi prvih opažanj ustvarili še drugi vzorec, v katerega smo vključili samo nestandardne tvite, katerih avtorji imajo v svojih besedilih največji delež tujejezičnih prvin (v nadaljevanju »fokusirani vzorec«).

Potek raziskave je na kratko predstavljen na sliki 2.



Slika 2: Načrtovan potek raziskave.

Že pred izdelavo raziskovalnega korpusa smo na podlagi teoretičnih izhodišč pripravili Smernice za označevanje (priloga A) in jih preizkusili na manjšem vzorcu naključnih tvitov, ki so vsebovali vsaj eno tujejezično prvino. Nato smo vsak preklop v splošnem in fokusiranem vzorcu, ki tvorita raziskovalni korpus, ročno označili na petih ravneh:

1. jezik (angleščina, nemščina, srbohrvaščina itd.),
2. vrsta preklopa (medstavčni, znotrajstavčni),
3. zapis (nepodomačen, delno podomačen, povsem podomačen),
4. morfologija (ni razvidna, razvidna iz končnice in/ali obrazila) in
5. besedna vrsta.

Kadar je bil preklop v heštegu, smo to označili z dodatno oznako.

Priprava in označevanje korpusa sta podrobneje predstavljena v razdelku 3.2. Analizo ročno označenih tvitov, ki vsebujejo vsaj en kodni preklop iz slovenščine v drug tuj jezik in so vključeni v korpus preklapov, smo opravili v dveh ločenih sklopih – kvantitativni (razdelek

3.3) je namenjen predvsem odkrivanju splošnih značilnosti kodnega preklapljanja in morebitnih slovničnih omejitev, s kvalitativnim (razdelek 3.4) pa smo poskušali ugotoviti, v kakšnih kontekstih in iz kakšnih razlogov ljudje uporabljajo tujejezične prvine.

3.1.1 Kvantitativna analiza

Ročno označen korpus preklopov smo izvozili v preglednico v formatu .xlsx, tako da smo jih lahko analizirali s programom Excel. Za preverjanje hipotez od 1) do 7) (gl. razdelek 1.2) smo s statistično analizo ugotavljali naslednje:

A. Splošne značilnosti preklapljanja

a. Pogostost preklopov

- Koliko tvitov v označenem korpusu je vsebovalo vsaj en kodni preklop?
- Koliko je bilo vseh kodnih preklopov?
- Kako natančno oz. ustrezno je avtomatsko označevanje tujih besed?
- Ali se pogostost preklapljanja razlikuje glede na standardnost besedila (L1, L2, L3)?

Že pri testnem označevanju manjšega vzorca tvitov, ki so vsebovali kakšno tujejezično besedo, smo ugotovili, da bomo zaradi nepopolnega avtomatskega označevanja po izbranih parametrih v vzorec dobili številne lažno pozitivne zadetke, torej tvite, v katerih dejansko ni preklopa. Odločili smo se, da bomo takšne tvite izločili iz raziskovalnega korpusa in nadaljnje analize. Poleg absolutnega števila tvitov s kodnim preklapljanjem in kodnih preklopov nas je zanimalo tudi, ali lahko v tako kratkih besedilih, kot so značilna za Twitter, ljudje večkrat preklopijo iz slovenščine v tuj jezik.

b. Uporabljeni jeziki

- V katere jezike najpogosteje preklapljajo slovenski tviteraši?
- Ali se razmerje jezikov spremeni glede na vrsto preklopa (znotraj- in medstavčni)?
- Ali obstajajo razlike glede uporabljenih jezikov med splošnim in fokusiranim vzorcem?

V naslednjem koraku smo prešteli oznake za jezik glede na druge lastnosti preklopov.

c. Dolžina preklopov

- Kakšno je razmerje med medstavčnimi ali znotrajstavčnimi preklopi?
- Koliko je znotraj- in medstavčnih preklopov z eno, dvema, tremi itd. besedami?
- Kako dolgi so preklopi v povprečju?
- Ali se dolžina preklopov razlikuje glede na vrsto preklopa, jezik in besedno vrsto?

Ob upoštevanju raziskav v drugih jezikih, kjer se je izkazalo, da so najpogostejši enobesedni preklopi (cf. Myers-Scotton 1997: 180), smo pričakovali, da bodo le-ti najštevilčnejši tudi v našem korpusu preklapov. Sledili jim bodo dvobesedni, trobesedni itd., pri čemer bo prevlada krajših preklapov bolj izrazita pri znotrajstavčnih preklapih, medstavčni pa bodo daljši, torej bodo vsebovali več besed. V zvezi z razlikami glede na jezik smo se odločili primerjati le tri najpogostejše jezike (angleščino, srbohrvaščino in nemščino), saj je število tvitov iz drugih jezikov premajhno, da bi lahko povlekli splošne zaključke. V primeru zanimivih odstopanj (npr. najdaljših preklapov) pa si bomo zadevne tvite ogledali nekoliko podrobneje.

B. Slovnične omejitve preklapljanja

- Katere besedne vrste so najpogostejše del kodnega preklopa?
- Ali obstajajo besedne vrste, ki ne morejo biti predmet kodnega preklapljanja?
- Ali se pogostost besednih vrst razlikuje glede na jezik?
- Katere besede se najpogostejše pojavljajo kot del preklopa?

V tem sklopu smo želeli preveriti, ali za kodno preklapljanje v slovenščini obstajajo kakšne omejitve glede tega, katere besedne vrste lahko nastopajo kot del preklopa. V teoriji ni konsenza o tem, ali splošne omejitve pri preklapljanju, ki izhajajo iz slovnične ureditve jezika, res obstajajo ali so morda vezane ne posamezne jezikovne kombinacije ali pa so uporabniki pri kombiniranju povsem prosti in lahko ustvarjalno prilagajajo rabo jezika svojim potrebam. Obseg naše raziskave seveda ne omogoča, da bi lahko podali izčrpen prispevek k tovrstnim diskusijam, bomo pa preizkusili nekatere vidike teorij predstavljenih v razdelku 2.4.6. Tudi tukaj bomo pri ugotavljanju razlik med jeziki upoštevali samo angleščino, srbohrvaščino in nemščino, poleg tega pa si bomo pri zanimivih opažanjih (npr. občutno večja zastopanost določene besedne vrste pri enem od jezikov) ogledali vsebino zadevnih tvitov in na podlagi kvalitativne analize poskusili utemeljiti, zakaj prihaja do takšnih razlik.

C. Integracija preklapov v slovenščino

- a. Ortografska integracija (zapis)
 - V koliko preklapih je zapis tuje prvine delno ali povsem podomačen?
 - Ali se stopnja podomačenosti razlikuje glede na jezik in besedno vrsto?

Glede na teorije, opisane v razdelku 2.4.1, po katerih je eden od možnih kriterijev za razlikovanje med kodnimi preklapi in prevzetimi besedami njihova integracija v matrični jezik (cf. Gardner-Chloros in Weston 2015, Poplack 2015), bomo v tem razdelku preverili hipotezo, da je večji delež kodnih preklapov nepodomačen, torej je jedro besede ohranilo isti zapis kot v

izvirniku (vključno s tujimi črkami). Kljub temu pa je treba pri interpretaciji podatkov upoštevati, da je posebej pri srbohrvaščini težavno ugotavljati, ali je zapis prilagojen slovenščini, zato smo pri ugotavljanju razlik glede na jezik upoštevali samo angleščino in nemščino. Pri tem smo pričakovali, da bodo nemški preklopi bolj podomačeni, saj so v slovenščini prisotni že dlje časa (številni germanizmi so pravzaprav že našli pot v slovenske priročnike, torej so po naših merilih prevzete besede in niti niso označeni v korpusu preklpov).

b. Morfološka integracija

- Pri kolikšnem deležu preklpov je razvidna morfološka integracija?
- Ali se deleži razlikujejo glede na jezik in besedno vrsto?

Morfološko integracijo smo lahko opazovali zgolj pri glagolih, pridevniki in samostalnikih. Pri samostalnikih smo že med pripravo smernic za označevanje predvideli težave zaradi ničtih končnic (npr. v imenovalniku), vendar smo se glede na število ostalih oznak odločili, da jih ne bomo označevali ločeno. Tudi v tem razdelku smo pri ugotavljanju razlik glede na jezik upoštevali le angleščino in nemščino, saj je srbohrvaščina tako kot slovenščina sintetični jezik, ki ima veliko istih oblik zaradi podobnosti med jeziki, zato bi v večini primerov težko presodili, ali je razvidna slovenska končnica ali pa so preklopi tvorjeni v skladu s slovnico vrinjenega jezika. Zaradi tega smo pri vseh tvitih z oznako za jezik »hbs« uporabili oznako za zapis »NK«.

c. Ortografska in morfološka integracija

- Koliko preklpov je delno/povsem prilagojenih?

Na koncu smo prešteli, koliko kodnih preklpov ima hkrati oznaki za delno/povsem podomačen zapis ter razvidno končnico in/ali obrazilo. Glede na omejitve, omenjene v točkah a. in b., smo pričakovali, da bo takšnih zadetkov precej malo, zato smo se odločili še za njihovo kvalitativno analizo, saj so med njimi glagoli in samostalniki, ki utegnejo z razširjeno rabo in sčasoma tudi z vključitvijo v slovarje postali prevzete besede.

V sklopu tega razdelka smo izdelali tudi seznam vseh besed iz kodnih preklpov in jih razvrstili po njihovi frekvenci. Na ta način bomo poskusili ugotoviti, ali obstajajo besede, ki so zaradi pogostosti rabe kandidatke za vključitev v slovensko besedišče (cf. Myers-Scotton 1997), seveda ob zavedanju, da imamo na voljo zgolj omejen korpus podatkov, zato naši zaključki ne morejo biti dokončni.

D. Razlike glede na spol

- Ali je preklapljanje bolj značilno za moške oz. ženske?
- Ali se moški in ženski kodni preklopi razlikujejo glede na zastopanost jezikov ali dolžino?

V literaturi do zdaj ni bilo veliko raziskav na temo preklapljanja in spola (cf. Gardner-Chloros 2009), vendar smo se glede na to, da imamo pri vseh tvitih v raziskovalnem korpusu označen spol avtorja, kot zanimivost odločili preveriti, ali morda obstajajo kakšne bistvene razlike v kodnem preklapljanju med moškimi in ženskami. Pri tem izhajamo iz hipoteze, da opaznih razlik ne bo mogoče najti.

3.1.2 Kvalitativna analiza

Priprave na kvalitativni del raziskave, s katerim smo preverjali hipoteze 8) do 10), so potekale že med označevanjem v programu WebAnno, in sicer smo si sproti beležili seznam potencialnih tematskih polj, iz katerih prihaja večje število preklpov (npr. IT, popularna kultura, kulturne specifike; cf. Tastan 2012), ter funkcije preklopa (npr. citati, ponovitve, nagovarjanje naslovnikov; cf. Gardner-Chloros 2009 in Androutsopoulos 2013b). Opazili smo tudi, da so daljši preklopi velikokrat pravzaprav idiomatične besedne zveze, ki jih uporabniki prevzemajo iz tujega jezika »kot celoto« (cf. Myers-Scotton 1997).

Po končanem označevanju smo vse tvite izvozili v datoteko .xlsx, iz katere smo lahko naredili seznam vseh tvitov, ki vsebujejo kodno preklapljanje. Nato smo si v preglednici v stolpce poleg posameznih tvitov zabeležili:

- tematsko/semantično polje (npr. IT, Twitter, šport)
- ali je kodni preklop idiom, metafora, stalna besedna zveza ali pregovor,
- druga opažanja, ki bi se lahko izkazala kot zanimiva.

Razlogov, zakaj v tem delu nismo vnaprej pripravili zaključenega nabora kategorij tem, funkcij ali motivacij za preklop, je več. Med pregledom relevantne literature smo ugotovili, da zaenkrat še ni uveljavljene univerzalne klasifikacije diskurznihih funkcij kodnega preklapljanja v računalniško posredovani komunikaciji, ki bi jo bilo mogoče uporabiti v poljubnih jezikovnih kombinacijah. Poleg tega je naš vzorec premajhen in preveč specifičen, da bi lahko na podlagi izsledkov sklepali o tem, kako pogosto se recimo kodno preklapljanje uporablja za doseganje posebnega poudarka.

Glavni razlog pa je predvsem praktičen. Pripisovanje motivacije je že na sploh težavno, pogosto

subjektivno, poleg tega pa ugotavljanje, ali je nekdo prej prikladal določeno besedo iz tujega/vrinjenega jezika namesto iz svojega/matričnega, ker jo uporablja tako pogosto, da je v njegovem besedišču pravzaprav že »prevzeta beseda«, ali pa jo je izbral, ker je želel (podzavestno?) zaznamovati svoj izrek, vsekakor presega okvire tega dela. Ob tem moramo upoštevati dejstvo, da je v označenem korpusu zbrana množica tvitov, ki so iztrgani iz konteksta, zato je še toliko težje presoditi, zakaj je nekdo uporabil tujejezično prvino.

Kljub temu smo se odločili predstaviti zanimivejše primere, ki smo jih opazili med označevanjem, in jih kvalitativno analizirati v omejenem obsegu. Pri tem smo poskušali vsaj deloma odgovoriti na naslednja vprašanja:

- Ali je mogoče kodnim preklpom pripisati določeno diskurzno funkcijo?
- V kakšnih kontekstih se pojavljajo preklopi?
- Ali se med preklopi pojavljajo frazeološke enote (idiomi, stalne besedne zveze itd.)?

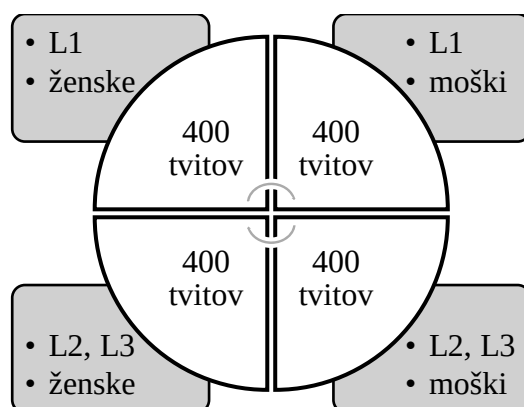
Naš cilj torej ni podati rezultatov, kako pogosto nastopajo kodni preklopi v točno določeni funkciji ali s katerega tematskega polja prihaja kakšen delež tujejezičnih prvin. Kljub temu so lahko naše ugotovitve podlaga za raziskovanje v to smer in bodo morda lahko služile kot izhodišče za podrobnejše študije kodnega preklapljanja v spletni slovenščini.

3.2 Priprava in označevanje korpusa preklpov

3.2.1 Priprava korpusa

Za potrebe ročnega označevanja tvitov oziroma kodnih preklpov smo najprej iz podkorpusa tvitov v korpusu Janes 0.4 pridobili 1600 naključnih tvitov, ki so vsebovali vsaj eno tujejezično besedo (oziroma besedo, ki je bila računalniško označena kot tuje z oznako Nj), od katerih jih je bila polovica standardnih (oznaka L1) in polovica nestandardnih (oznaki L2 in L3). Dodaten pogoj je bil, da gre za tvite uporabnikov, ki imajo vsaj 1000 tvitov, poleg tega pa so bili uravnoteženi tudi po spolu, tako da je bila polovica avtorjev ženskih, polovica pa moških uporabnikov. V raziskavi smo upoštevali samo tvite zasebnih uporabnikov.

Prvi vzorec je bil sestavljen iz štirih delov, ki so ponazorjeni na sliki 3.

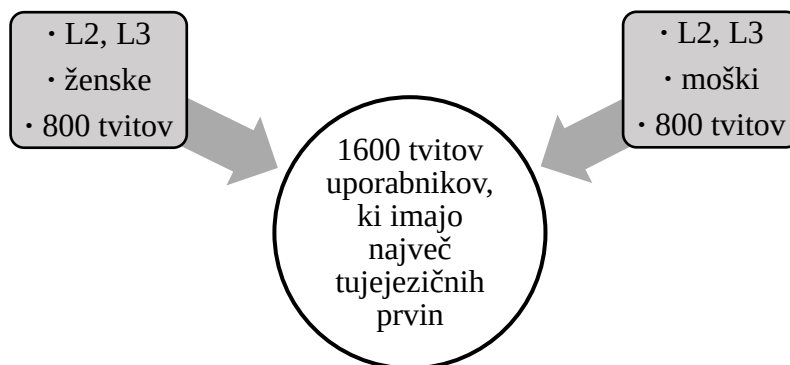


Slika 3: Značilnosti tvitov, vključenih v prvi raziskovalni vzorec.

Med označevanjem tega prvega, »splošnega« vzorca se je izkazalo, da je zelo veliko lažno pozitivnih zadetkov oziroma tvitov, v katerih ni kodnega preklopa. Prvi razlog za to je bil, da v raziskavi nismo upoštevali tvitov, ki so v celoti v tujem jeziku oz. brez slovenskih besed. Drugi razlog izhaja iz pomanjkljivosti samodejnega oblikoskladenjskega označevanja, ki ni zelo natančno, zato je bila oznaka Nj pogosto napačno pripisana tudi slovenskim besedam. Vse takšne tvite smo ročno izločili iz nadaljnje obdelave, tako da smo jim v aplikaciji WebAnno na začetek dodali oznako \$0 in jih nato izbrisali pred izvozom podatkov v format .xlsx.

Na podlagi teh opažanj smo se odločili razširiti raziskavo še z drugim, »fokusiranim« vzorcem. Za izhodišče smo ponovno vzeli 1600 tvitov, polovico ženskih in polovico moških. Med označevanjem preklpov v prvem vzorcu smo ugotovili, da se tviti precej pogosteje pojavljajo v nestandardnih besedilih. Ker smo želeli z drugim vzorcem pridobiti večje število relevantnih tvitov (s »pravimi« kodnimi preklpi, ne lažno pozitivnimi zadetki), smo se odločili, da morajo imeti vsi tviti v tem vzorcu oznako nestandardno (L2, L3). Dodatno smo želeli rezultate

izboljšati tako, da smo vanj vključili po 20 naključnih tвитov od 80 avtorjev, ki imajo največ tujih besed (Nj) v korpusu Janes 0.4. Na ta način smo dobili torej vzorec tвитov oseb, prikazan na sliki 4, za katere predvidevamo, da pogosto uporabljajo kodno preklapljanje.

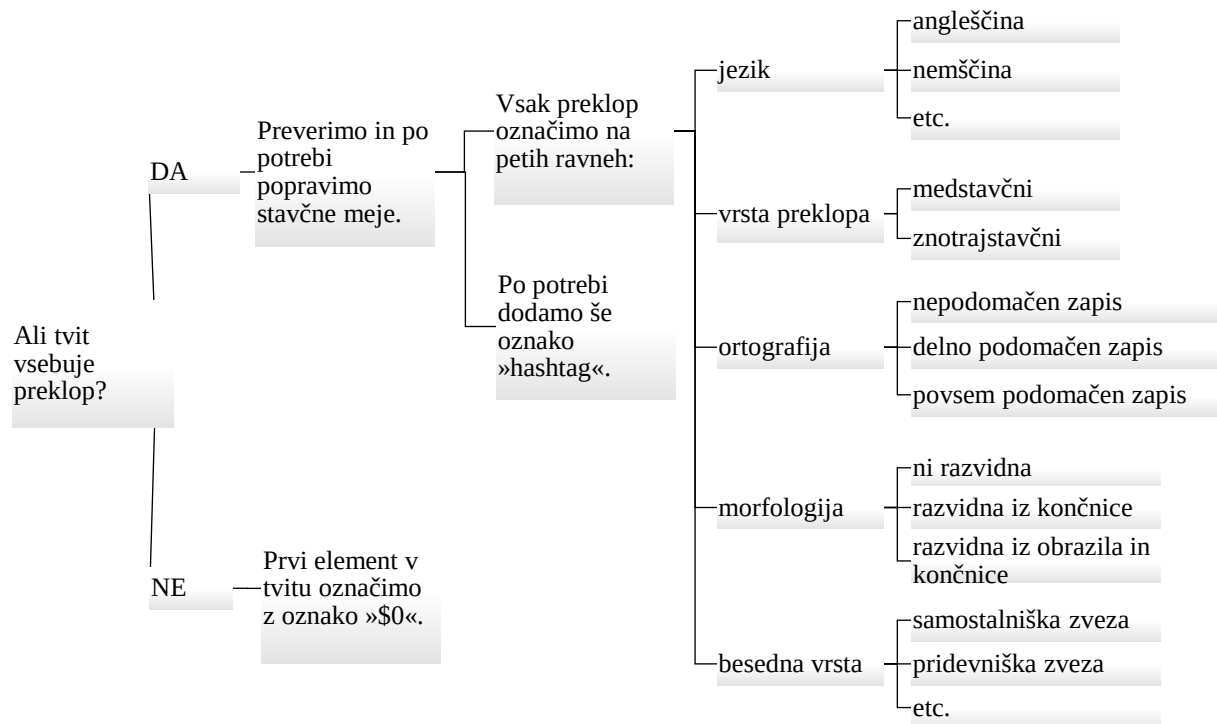


Slika 4: Značilnosti tвитov, vključenih v drugi raziskovalni vzorec.

Pri analizi smo zaradi tega pristopa določene značilnosti opazovali samo na prvem vzorcu (npr. primerjave glede na (ne)standardnost). Za večino ostalih lastnosti pa smo najprej prešteli število posameznih oznak za vsak vzorec posebej – kjer med njima ni bilo bistvenih razlik, smo rezultate predstavili skupno, tako da smo sešteli vse oznake za določeno kategorijo.

3.2.2 Označevanje preklapov

Označevanje kodnih preklapov je potekalo na odprtokodni platformi WebAnno, ki je namenjena ročni anotaciji korpusov (Erjavec et al. 2016: 2). Zaporedje posameznih korakov pri označevanju je prikazano na sliki 5.



Slika 5: Shema poteka označevanja kodnih preklompov v programu WebAnno.

Pred začetkom dela smo na podlagi manjšega vzorca testnih primerov iz korpusa Janes 0.4 pripravili smernice za označevanje, ki so vključene v prilogi A, zato jih spodaj povzemamo zgolj v strnjeni obliki.

A. Izločitev nerelevantnih tvitov

Iz raziskovalnega korpusa smo izločili tvite, ki so v celoti v tujem jeziku ali več tujih jezikih, saj nas v raziskavi zanima kodno preklapljanje v slovenščini. Izločili smo tudi avtomatsko generirane in popolnoma nerazumljive tvite. Poleg tega nismo označevali tujejezičnih prvin, ki so se v tvitu pojavile kot del retvita (cf. Page et al. 2014).

B. Označevanje stavčnih mej

Najprej smo v vsakem tvitu preverili, ali je avtomatska stavčna segmentacija pravilna. Pri tem smo se ravnali po *Smernicah za označevanje korpusa slovenskih tvitov JANES* (v nadaljevanju: »Osnovne smernice«). Pri tem je ključno izpostaviti, da »stavek« uporabljamo kot najmanjšo samostojno enoto jezikovnega sporočila, ki bi se v standardni slovenistični začela z veliko začetnico in končala s končnim ločilom (»poved« v slovenistični terminologiji). Težavni primeri so pojasnjeni v prilogi A.

C. Oznake na prvi ravni: jezik in vrsta kodnega preklopa

a. Določanje jezika

Kot slovenske smo upoštevali besede, ki so vključene v Slovar slovenskega knjižnega jezika (SSKJ), Slovenski pravopis (PS) ali Sloleks (cf. Deuchar 2006 in Rosenberg 2013). V primeru dvoma, ali je beseda v tujejezičnem kontekstu slovenska ali ne, jo štejemo kot del prevladujočega jezika v tvitu. Smiselno podobno velja za besedne zveze, kjer je en element mogoče šteti kot slovensko besedo, vendar je treba celotno zvezo šteti kot zaključeno enoto (npr. v »buffalo burger« je tudi »burger« angleška beseda).

Posebnost je pogosto uporabljena latinska okrajšava za *versus* (vs, vs., v.), ki smo jo označevali kot angleško, saj predpostavljamo, da jo uporabniki prevzamejo iz angleških besedil. To je bilo razvidno tudi iz konteksta (npr. »V soboto se oglasim s tekme Crystal Palace vs. Chelsea«).

V smernice smo vključili tudi tujejezične prvine in primere, ki jih nismo označevali kot preklomp besede, ki jim ni mogoče določiti izvora, neologizme, avtomatske tvite oziroma dele tvitov, naslove pesmi, člankov, spletnih strani, interference na nivoju skladnje, tujejezične prvine v citiranem tvitu ter slovenske besede, zapisane s tujo črko.

b. Vrsta preklopa

Preklope smo označili bodisi kot znotrajstavčne bodisi kot medstavčne (cf. Myers-Scotton 1997, Poplack 1980, Appel in Muysken 2005). Medstavčni so tisti preklopi, kjer je celoten stavek (ali več stavkov) v istem tujem jeziku.

D. Oznake na drugi ravni: zapis (ortografija), morfologija in besedna vrsta

a. Posebnost – večbesedni preklopi

Kot eno enoto, kateri smo pripisali po eno skupno oznako za zapis, morfologijo in besedno vrsto, smo obravnavali samostalniške zveze s prilastki, pridevniške zveze s prilastki, predložne zveze, sklope ponovljenih besed, ključnike ter stavke in daljše sklope besed, ki jim ni bilo mogoče določiti ene besedne vrste.

Znotrajstavčne preklope, ločene s stavčno mejo, smo označili vsakega zase, čeprav bi lahko rekli, da gre za enkratni preklomp v tuj jezik. Pri medstavčnih preklompah pa stavčne meje ne upoštevamo, torej več zaporednih medstavčnih preklompov označimo kot en preklomp.

b. Zapis

Pri označevanju ortografije smo uporabili tri kategorije: podomačen, delno podomačen in

nepodomačen. Zapis smo označevali glede na osnovno obliko oziroma jedro besed, pri čemer nismo upoštevali obrazil ali končnic. Nepodomačeni so kodni preklopi, kjer je ohranjen isti zapis, kot je v izvirnem jeziku. Delno podomačeni so tisti, kjer je opazna sprememba izvirnega zapisa, npr. z izpuščenim končnim e-jem ali dvojnimi soglasnikom, vendar še vedno vsebujejo tuje elemente oziroma ne odražajo v celoti »slovenske« izgovorjave. Kot delno podomačene smo označili tudi mešane primere, kjer je del besedne zveze povsem prilagojen, del pa je zapisan izvirno.

c. Morfologija

Za označevanje morfoloških značilnosti smo uporabili štiri oznake, in sicer glede na to, ali je razvidna slovenska končnica (spreganje oziroma sklanjanje), dodano obrazilo (besedotvorje) ali oboje. Večbesedne zveze smo označili z eno oznako glede na njihovo jedro.

d. Besedna vrsta

Pri določanju besednih vrst smo smiselno upoštevali Osnovne smernice in pravila za označevanje večbesednih preklpov. Daljše besedne zveze oziroma cele stavke smo označili z oznako »neuvrščeno«. Pri ugotavljanju, katere besedne vrste so najpogostejše predmet preklapljanja v slovenščini, smo tako upoštevali le krajše besedne zveze. To nam zadostuje za sklepanje o veljavnosti slovničnih omejitev, predvsem omejitve glede možnosti preklapljanja zaprtih besednih vrst (cf. Joshi 1982 in Myers-Scotton 1997) – daljše otoke vrinjenega jezika namreč ureja slovnica tega jezika. Podrobnejše smernice glede posameznih besednih vrst so v prilogi A.

E. Dodatna oznaka: hashtag

Dodali smo jo samo, kadar se tujejezični element pojavi kot hešteg ali del heštega.

Na sliki 6 je posnetek zaslona, ki prikazuje okolje v programu WebAnno in pet označenih tvitov, od katerih trije vsebujejo skupno štiri kodne preklope, dva pa sta nerelevantna za raziskavo in sta bila izločena (prvi je namreč v celoti slovenski, drugi pa angleški).

The screenshot displays the WebAnno web application interface. The top navigation bar includes the 'Annotation' title, 'WebAnno | Home' link, and user information 'User: spela reher | Log out'. Below this is a toolbar with 'Document' (Open, Prev, Next, Export, Settings) and 'Page' (First, Prev, Go to 1, Next, Last) sections. Further right are 'Script' (LTR/RTL), 'Help' (Guidelines), and 'Workitow' (Done) buttons. The status bar at the bottom indicates 'JANES-Preklop 2/FslvL23.ok-001.tsv' and 'showing 1-5 of 40 sentences'.

The main annotation area shows five tweets, each with linguistic tags above the text:

- Tweet 1:** @Jure_Bajic nope ... prvic vidim to sranje ... do zdaj je delalo ...
 Tags: L, NP, NR, ANG, MS, \$.
- Tweet 2:** Ups , I did it again ... zdej je rekla , bo pa sam še štrik pel , nč več free like a birdy #Meggie http://t.co/6tm1WqcU2j
 Tags: N, DP, NR, ANG, MS, P, NP, NR, ANG, ZS, \$.
- Tweet 3:** @LazarjevPolzek če ga sučeš sam okol pasu , ti nardi obliko ! Če daš musko do konca in plešeš not , je kul anger management ;)
 Tags: S, NP, NR, ANG, ZS, \$.
- Tweet 4:** @ninagaspari raj si eno njivo vzem v najem , bo cenejš ;) @LanaDakic
 Tags: \$.
- Tweet 5:** @Jure_Bajic @mashy3 lol
 Tags: \$.

Slika 6: Posnetek zaslona s petimi tviti po končanem označevanju.

3.3 Kvantitativna analiza

Podatke iz ročno označenega korpusa smo izvozili v preglednico v formatu .xlsx, v katero niso bili vključeni tisti tviti, ki smo jih v programu WebAnno označili z \$0. Tako smo lahko relevantne tvite, ki tvorijo korpus preklpov, analizirali s programom Excel, v katerem smo prešteli ustrezne oznake. V nadaljevanju so prikazani končni rezultati, nekatere podrobnejše preglednice pa so dodane v prilogah.

3.3.1 Splošne značilnosti preklapljanja

3.3.1.1 *Pogostost preklpov*

Po izvozu smo najprej prešteli število tvitov, ki vsebujejo vsaj en kodni preklop. Ker smo že med označevanjem splošnega vzorca ugotovili, da je v raziskovalnem korpusu veliko šuma, smo glede na to prilagodili parametre za drugi, fokusirani vzorec, da bi dobili več pravih preklpov. Iz tabele 1 je razvidno, da nam je to tudi uspelo, saj se je delež relevantnih tvitov povečal z 26 % na 43 %. Kljub temu je zaradi že omenjenih razlogov (v celoti tujejezični tviti, neupoštevanje povezav na skladbe, članke ipd.) delež tvitov, ki smo jih morali izbrisati ročno, ostal dokaj velik. Na koncu je ročno označeni korpus tvitov vseboval 1100 tvitov oz. približno tretjino izhodiščnih tvitov v raziskovalnem korpusu.

	<i>Splošni vzorec</i>	<i>Fokusirani vzorec</i>	<i>Skupaj</i>
Število tvitov v raziskovalnem korpusu	1600	1600	3200
Število tvitov z vsaj enim preklpom	416	684	1100
Delež relevantnih tvitov glede na raziskovalni korpus	26 %	43 %	34 %

Tabela 1: Število in delež tvitov z vsaj enim kodnim preklpom.

Kot naslednje nas je zanimalo, koliko je bilo vseh kodnih preklpov v korpusu preklpov. Številke so predstavljene v tabeli 2. Ne preseneča, da je preklpov več kot tvitov (skupaj 1381), saj lahko posameznik v tvitu preklopi več kot enkrat. V povprečju so tako tviti vsebovali 1,26 preklopa na tvit. V šestih tvitih so bili uporabljeni po štirje kodni preklopi, kar je tudi najvišje število preklpov v posameznem tvitu (primer iii.).

[iii.] uf, good luck. jst sm red velvet kapkejke jedla. biskvit, pofarban na rdeč. ne štekam hajpa razn tega da je pretty.

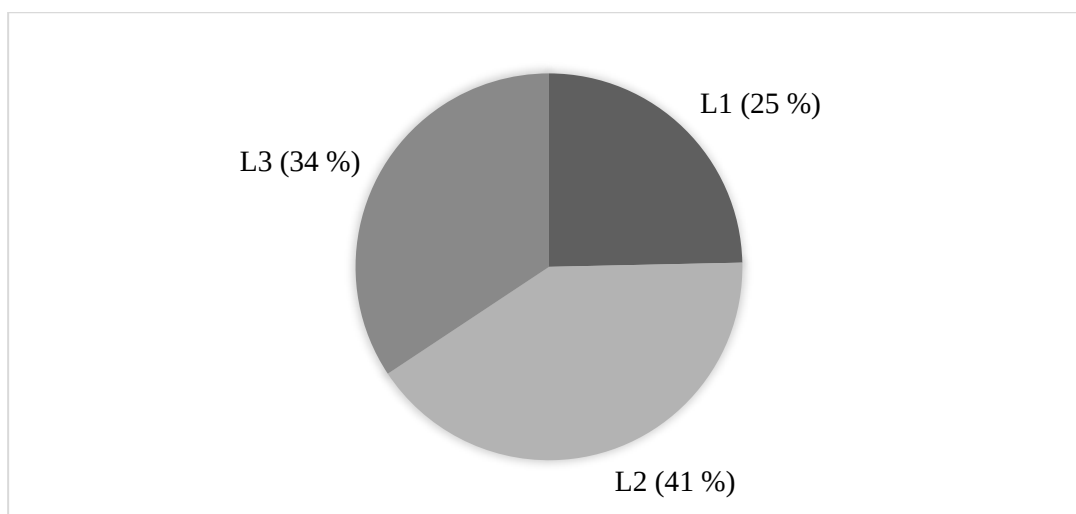
V skladu s pričakovanji je delež znotrajstavčnih preklpov višji od medstavčnih, kar je prvi korak k potrditvi hipoteze, da ljudje pogosteje prevzemajo posamezne besede kot cele stavke, čemur se bomo podrobneje posvetili še v razdelku 3.3.1.3.

	<i>Splošni vzorec</i>	<i>Fokusirani vzorec</i>	<i>Skupaj</i>
Skupno število preklpov	507	874	1381
Število (delež) medstavčnih preklpov	160 (32 %)	298 (34 %)	458 (33 %)
Število (delež) znotrajstavčnih preklpov	347 (68 %)	576 (66 %)	923 (67 %)
Povprečno število preklpov na tvit	1,22	1,28	1,26

Tabela 2: Število in delež preklpov v korpusu preklpov.

Na vprašanje, kako primerno je avtomatsko označevanje tujih besed za sklepanje o pogostosti preklapljanja na slovenskem Twitterju, na podlagi razpoložljivih podatkov ni mogoče podati zanesljivega odgovora. V korpusu Janes 0.4 je v tvitih, ki so označeni kot slovenski, tujejezičnih besed, ki se začnejo z malo začetnico (s tem se izognemo npr. naslovom pesmi in filmov), 1.404.792 oziroma 13.122,4 na milijon. Tudi če bi na podlagi naših izsledkov ocenili, da gre zgolj pri tretjini tvitov z avtomatsko označeno tujo besedo dejansko za kodni preklp, bi lahko rekli, da ne gre za zanemarljiv pojav. Vendarle pa je naš korpus relativno majhen in se ga ne da zanesljivo posplošiti, še posebej ob upoštevanju dejstva, da smo pri izbiranju tvitov za drugi, fokusirani vzorec določili take kriterije, po katerih naj bi bilo preklpov v vzorcu čim več. Poleg tega ostaja odprto vprašanje, koliko je kodnih preklpov, ki jih avtomatski označevalnik označi kot slovenske besede.

Pri analizi samo prvega vzorca, ki je vseboval 507 preklpov, smo ugotovili, da se pogostost preklapljanja razlikuje glede na standardnost besedila (L1, L2, L3), in sicer je v bolj standardnih besedilih delež kodnega preklapljanja nižji. Z grafa 3 je razvidno, so tri četrtine vseh preklpov izhajale iz tvitov z oznako standardnosti L2 ali L3, le četrtina pa jih je bila označenih z L1.



Graf 3: Deleži preklpov glede na standardnost besedila v splošnem vzorcu (N = 507).

Standardnost besedil je bila označena avtomatsko, zato smo si na koncu ogledali, ali je morda

prišlo do šuma zaradi nezanesljivosti avtomatskega označevanja. Izkazalo se je, da so deli tvitov, ki so zapisani v slovenščini, dejansko večinoma slovnično pravilni, poleg tega pa vsebujejo določen tujejezični dodatek. Izkaže se, da je preklon včasih v obliki heštega, ki je dodan na konec slovenskega besedila (primer iv.):

[iv.] Ob navijanju komentatorjev za odpoved veleslaloma in druge serije poletov vsaj naši športniki dajejo vzgled. Bravo #orli #fairplay.

Kot je razvidno iz tabele 3, je v tvitih, označenih z najvišjo stopnjo standardnosti, delež heštegov največji.

	<i>Število heštegov</i>	<i>Število preklonov</i>	<i>Delež heštegov</i>
L 1 – splošni vzorec	15	125	12,0 %
L2, L3 – splošni vzorec	10	382	2,6 %
L2, L3 – fokusirani vzorec	23	874	2,6 %

Tabela 3: Število in delež heštegov v kodnih preklonih.

3.3.1.2 Uporabljeni jeziki

Najprej smo prešteli oznake za jezik za vsak vzorec posebej, vendar so bile razlike med njima zanemarljive, zato so v tabeli 4 predstavljeni le skupni rezultati. Potrdili smo hipotezo, da angleščina z 90-odstotnim deležem občutno prevladuje kot jezik, v katerega preklaplajo slovenski tviteraši. Sledita ji srbohrvaščina (4,6 %) in nemščina (3,3 %), ostalih jezikov pa je bolj za vzorec (skupaj 2,1 %), vendar je nabor dokaj raznolik – skupno smo namreč zaznali kar 9 različnih jezikov, pri čemer nismo razlikovali med bosanščino, hrvaščino in srbsčino.

	<i>Število preklonov</i>	<i>Delež</i>
angleščina	1244	90,1 %
srbohrvaščina	63	4,6 %
nemščina	45	3,3 %
latinščina	9	0,7 %
španščina	9	0,7 %
francoščina	6	0,4 %
italijanščina	3	0,2 %
arabščina	1	0,1 %
poljščina	1	0,1 %
skupaj	1381	100 %

Tabela 4: Število in delež preklonov po jezikih v korpusu preklonov.

Nato smo preverili, ali se razmerje med jeziki spremeni glede na vrsto preklopa. Podatki za cel korpus preklonov so navedeni v tabeli 5. Razlike niso zelo velike, vendar pa je delež angleščine med medstavčnimi prekloni še višji (95 %), manj pa je srbohrvaških (2,8 %) in nemških

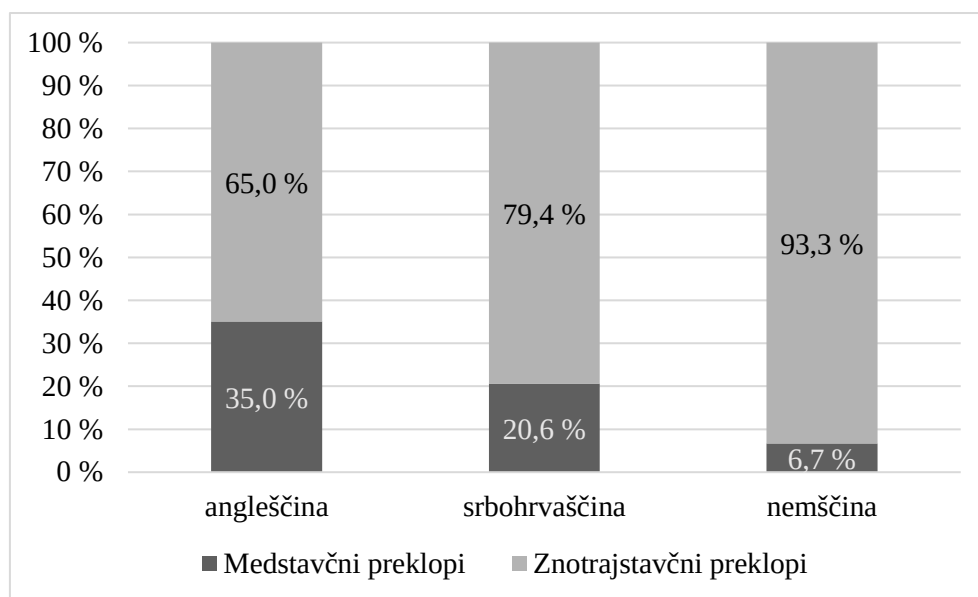
preklopov (0,7 % oziroma le 3 primeri), med ostalimi pa sta dva španska preklopa, dva latinska (oba »Mea culpa«) ter po eden iz italijanščine in arabščine (kar skupaj predstavlja 1,3 %).

Nasprotno je pri znotrajstavčnih preklapih razpršenost jezikov nekoliko večja: delež angleščine je nižji (87,5 %), več pa je srbohrvaščine (5,4 %), nemščine (4,6 %) in ostalih jezikov (skupaj 23 primerov oziroma 2,5 %). Če se že na tem mestu navežemo na kvalitativno analizo (gl. razdelek 3.4), lahko po hitrem preletu preklapov iz teh manj prisotnih jezikov ugotovimo, da gre predvsem za uveljavljene formulaične izraze (npr. »bon voyage«, »žnesekua«, »buenas noches«) ali kulturno specifične pojme (npr. »prosecco«, »el clasico«, »bachata«). Celoten seznam je v prilogi B.

	<i>Medstavčni</i>	<i>Znotrajstavčni</i>	<i>Vsi preklopi</i>
angleščina	95,2 %	87,5 %	90,1 %
srbohrvaščina	2,8 %	5,4 %	4,6 %
nemščina	0,7 %	4,6 %	3,3 %
latinščina	0,4 %	0,8 %	0,7 %
španščina	0,4 %	0,7 %	0,7 %
francoščina	0,0 %	0,8 %	0,4 %
italijanščina	0,2 %	0,2 %	0,2 %
arabščina	0,2 %	0,0 %	0,1 %
poljščina	0,0 %	0,1 %	0,1 %
skupaj	100 %	100 %	100 %

Tabela 5: Deleži jezikov glede na vrsto preklopa.

Nazadnje smo za najpogostejše tri jezike preverili, kakšno je razmerje med medstavčnimi in znotrajstavčnimi preklapi. Pri angleščini znaša razmerje približno 1 : 2, pri srbohrvaščini približno 1 : 4, pri nemščini pa kar 1 : 14 v korist znotrajstavčnih preklapov, kar je razvidno tudi na grafu 4. Na podlagi tega, da so preklopi v nemščino skoraj izključno znotrajstavčni, smo oblikovali dodatno hipotezo, da ljudje uporabljajo posamezne nemške besede oziroma germanizme, ki so zaradi dolgotrajne prisotnosti bolj integrirane v slovenščino, ne poslužujejo pa se nemških stalnih besednih zvez ali idiomov. Na to hipotezo bomo pozorni tudi pri naslednjih sklopih, kjer bomo zbrali več podatkov za končne zaključke.



Graf 4: Deleži medstavčnih in znotrajstavčnih preklopov v posameznem jeziku.

3.3.1.3 Dolžina preklopov

Že iz tabele 2 je jasno, da je znotrajstavčnih preklopov več kot medstavčnih, in sicer v razmerju 1 : 2. Za izhodišče smo predpostavili, da so znotrajstavčni preklopi krajši, saj se pri le-teh preklopi samo del stavka, pri medstavčnih pa (najmanj) en stavek, vsaj v povprečju pa bo del stavka krajši od celih stavkov, čeprav so seveda tudi ti lahko zgolj enobesedni. Na podlagi tega lahko postavimo hipotezo, da je v korpusu preklopov več krajših preklopov kot daljših.

Za potrditev te hipoteze smo v naslednjem koraku prešteli, koliko je enobesednih, dvobesednih, trobesednih itd. preklopov (glede na vrsto preklopa in skupaj). Razpon dolžine, merjene v besedah, je znašal od 1 do 13, vendar v nadaljevanju v tabeli 6 predstavljamo rezultate le za eno- do sedembesedne preklope. Daljših je bilo namreč zelo malo (vseh skupaj 24 primerov), kot zanimivost pa jih navajamo v prilogi B.

	<i>Znotrajstavčni</i>	<i>Medstavčni</i>	<i>Vsi preklopi</i>
1-besedni	61,0 %	29,5 %	50,5 %
2-besedni	21,9 %	21,6 %	21,8 %
3-besedni	9,4 %	17,2 %	12,0 %
4-besedni	3,9 %	14,4 %	7,4 %
5-besedni	1,2 %	7,0 %	3,1 %
6-besedni	1,4 %	3,3 %	2,0 %
7-besedni	0,7 %	2,8 %	1,4 %
8- do 13-besedni	0,5 %	4,1 %	1,7 %
skupaj	100 %	100 %	100 %

Tabela 6: Deleži preklopov glede na dolžino in vrsto preklopa.

Iz zgornje preglednice je razvidno, da so enobesedni preklopi daleč najštevilčnejši in predstavljajo polovico vseh preklpov. Drugi največji delež predstavljajo dvobesedni preklopi. Na podlagi teh podatkov lahko potrdimo hipotezo, da govorci najpogosteje uporabijo enobesedne tujejezične elemente, kar ni presenetljivo, saj lahko posamezne besede ali krajše besedne zveze v svojem diskurzu uporabijo tudi ljudje, ki tujega (vrinjenega) jezika niti ne obvladajo. Ta ugotovitev nakazuje tudi slabost raziskovalnega pristopa, v katerem se avtorji odločijo, da je treba vse enobesedne elemente šteti za izposojenke in torej niso predmet preučevanja kodnega preklapljanja, saj se na ta način gotovo izgubi pomemben del podatkov.

Zbrane podatke smo analizirali še z drugega vidika, in sicer tako, da smo izračunali povprečne dolžine znotraj- in medstavčnih preklpov, navedene v tabeli 7. Izkaže se, da so medstavčni preklopi v povprečju 70 % daljši od znotrajstavčnih, s čimer smo potrdili izhodiščno hipotezo.

	<i>Povprečno število besed</i>
Znotrajstavčni preklopi	1,73
Medstavčni preklopi	2,94
Vsi preklopi	2,13

Tabela 7: Povprečna dolžina preklpov glede na vrsto preklopa.

Za ugotavljanje, ali se dolžina preklpov razlikuje glede na jezik, smo prešteli število tvitov, dolgih od ene do sedem besed, in sicer za tri najpogostejše jezike. Deleži tvitov določene dolžine glede na vse tvite v posameznem jeziku so predstavljeni v tabeli 8.

	<i>angleščina</i>	<i>srbohrvaščina</i>	<i>nemščina</i>
1-besedni	49,8 %	55,6 %	80,0 %
2-besedni	21,9 %	14,3 %	11,1 %
3-besedni	12,2 %	9,5 %	4,4 %
4-besedni	7,5 %	12,7 %	2,2 %
5-besedni	3,2 %	4,8 %	0,0 %
6-besedni	2,2 %	1,6 %	0,0 %
7-besedni	1,4 %	0,0 %	2,2 %
8- do 13-besedni	1,8 %	1,6 %	0 %
skupaj	100 %	100 %	100 %

Tabela 8: Deleži preklpov glede na dolžino in jezik preklopa.

Pri nemščini smo podobno kot zgoraj, kjer smo zaznali občutno višji delež znotrajstavčnih preklpov, ugotovili, da izstopa po večjem deležu enobesednih preklpov – 80 %, pri angleščini in srbohrvaščini pa delež znaša okrog ene polovice. Zanimiv je tudi nekoliko višji delež štiribesednih preklpov pri srbohrvaščini – po ogledu zadevnih primerov smo ugotovili, da gre za stalne/idiomatične besedne zveze, zato smo tudi te vključili v prilogo B. Glede na podatke v

tabeli 8 pa tudi ne presenečajo izračunane dolžine preklopov v tabeli 9, ki kažejo, da so nemški preklopi v povprečju krajši od povprečja oziroma od angleških in srbohrvaških.

	<i>Povprečno število besed</i>
angleščina	2,16
srbohrvaščina	2,10
nemščina	1,40

Tabela 9: Povprečna dolžina preklopov glede na jezik.

V zvezi z dolžino preklopov nas je zanimalo tudi to, ali oziroma kako se razlikuje glede na besedno vrsto. V tabeli 10 je predstavljeno, kolikšen delež so predstavljali eno- in večbesedni preklopi znotraj posameznih besednih vrst, ki so razvrščene po povprečni dolžini od najdaljših do najkrajših. Ne preseneča, da so v povprečju najdaljši »neuvrščeni« preklopi, saj gre pravzaprav za daljše besedne zveze ali cele stavke, ki jim ni bilo mogoče določiti ene same besedne vrste (primer v.); dobrih 65 % od le-teh so pravzaprav medstavčni preklopi, ki so seveda daljši (primer vi.)

[v.] »a te je gasilec pošpricu v oči? :O ti mu kr povej, da po faci nima kej, he has to treat you with respect :)
[vi.] i thought some Joe guy killed Bruce's parents? :) pa še to nismo zihr. Me zanima kako bojo v #Gotham to naredl :)

Po dolžini nato sledijo predložne zveze; te so sicer lahko samo enobesedne (npr. »by« ali »via«), večinoma pa so sestavljene iz več besed (npr. »for your information«, »in the mood« ali »mit extra Creme«). Naslednje so samostalniške besedne zveze, kar ne preseneča, saj samostalnike pogosto spremljajo levi ali desni prilastki (npr. »mission impossible«, »international no bra day« ali »Happy news«). Tudi prislovne zveze so v povprečju večbesedne (npr. »all summer long« ali »Kind of«), vendar pa je tukaj delež enobesednih že več kot 60-odstoten. Tudi med samostalniki je več kot 55 % enobesednih.

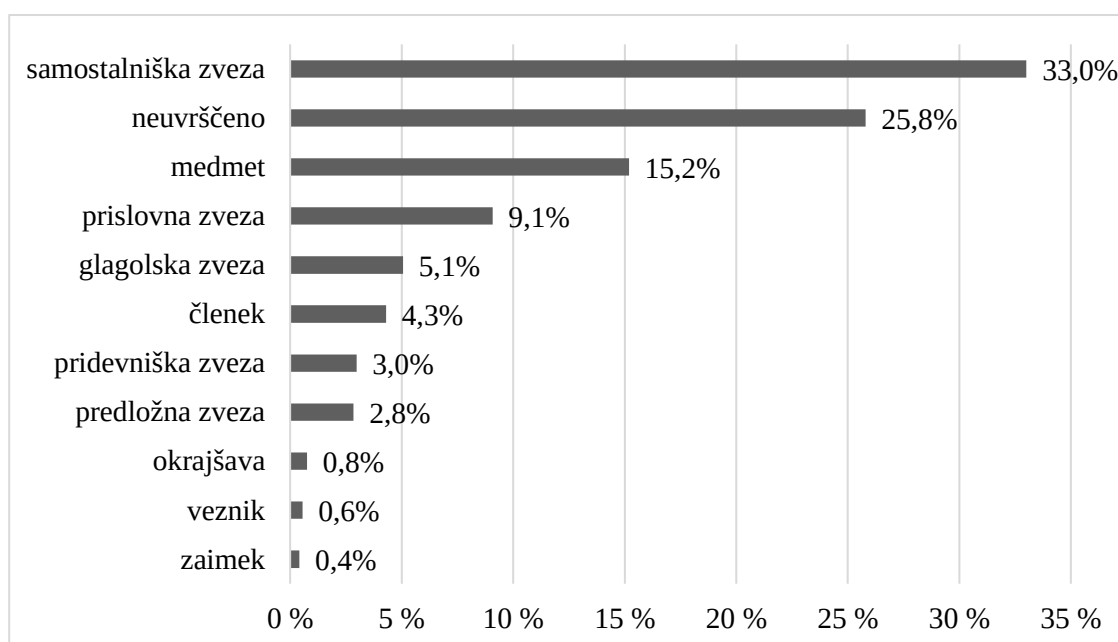
	<i>1-besedni.</i>	<i>2-besedni</i>	<i>3-besedni</i>	<i>4-besedni</i>	<i>5-besedni</i>	<i>6- do 13- besedni</i>	<i>Skupaj</i>	<i>Povprečna dolžina</i>
neuvrščeno	1,1 %	21,5 %	30,2 %	23,7 %	11,2 %	12,3 %	100 %	3,738
predložna zveza	32,6 %	37,2 %	20,9 %	7,0 %	2,3 %	0 %	100 %	2,093
samostalniška zveza	55,7 %	32,8 %	6,7 %	3,2 %	0,8 %	0 %	100 %	1,634
prislovna zveza	61,0 %	31,7 %	4,9 %	2,4 %	0 %	0,8 %	100 %	1,488
veznik	87,5 %	0 %	12,5 %	0 %	0 %	0 %	100 %	1,250
pridevniška zveza	80,9 %	14,5 %	3,8 %	0,8 %	0 %	0 %	100 %	1,244
medmet	88,1 %	9,1 %	2,3 %	0,5 %	0 %	0 %	100 %	1,151
glagolska zveza	86,3 %	13,7 %	0 %	0 %	0 %	0 %	100 %	1,137
členek	96,8 %	3,2 %	0 %	0 %	0 %	0 %	100 %	1,032
okrajšava	100 %	0 %	0 %	0 %	0 %	0 %	100 %	1,000
zaimek	100 %	0 %	0 %	0 %	0 %	0 %	100 %	1,000

Tabela 10: Deleži preklapov glede na dolžino in besedno vrsto.

V spodnji polovici preglednice so besedne vrste, kjer je velika večina primerov (več kot 80 %) enobesednih. Med njimi je znaten delež dvobesednih še pri pridevniki (npr. »Pretty cool« ali »custom made«), glagolih (»come on« ali »se šunja«) in medmetih (»oh yeaaah« ali »damn right«). Za veznike, členke, okrajšave in zaimke pa lahko rečemo, da so vedno enobesedni – za odstopanje v deležih in povprečni dolžini so pri veznikih in členkih pravzaprav krivi zgolj trije preklopi, kjer gre za ponovitev ene besede (»and and AND«, »jp jp« in »jp, jp«).

3.3.2 Slovnice omejitve preklapljanja

V skladu z načrtom raziskave smo najprej prešteli število preklapov glede na posamezno besedno vrsto v celotnem korpusu preklapov. Podatki so predstavljeni na grafu 5. Po pričakovanjih se je izkazalo, da kot preklopi najpogosteje nastopajo samostalniki oziroma samostalniške besedne zveze (33 %). Na drugem mestu so besedne zveze oz. stavki, ki jim ni bilo mogoče določiti besedne vrste. Kot smo ugotovili v predhodnem razdelku, gre za daljše fraze, pravzaprav »otoke vrinjenega jezika«, ki so tvorjeni v skladu s slovnico tujega jezika.



Graf 5: Deleži preklpov glede na besedno vrsto.

Na tretjem mestu so medmeti (15 %), ki so nas nekoliko presenetili, zato smo se odločili, da jim posvetimo nekaj več pozornosti. V prilogi C je tako vključen seznam medmetov, razvrščenih po jeziku in abecedi, iz katerega je razvidno, da je za velik delež medmetov pravzaprav zaslužno manjše število besed, ki se večkrat ponovijo, med katerimi izstopajo t.i. klepetalniške oz. spletne krajšave: »lol« (64 ponovitev, kar predstavlja skoraj tretjino vseh medmetov), »btw« (11), »omg/omfg« (10), »wtf« (4), »imho« (2). Precej ponovitev imajo tudi izrazi »thanks« (14), »sorry« (9) in »please« (6) v različnih nestandardnih zapisih (v oklepajih je zapisano število ponovitev v označenem korpusu).

V nasprotju s pričakovanji je bil delež pridevniških besednih zvez (9 %) in glagolov (5 %) nekoliko nižji. Sklepamo lahko, da je v slovenščino težje vriniti tuj glagol, saj ga je treba ustrezno spregati v skladu s pravili slovnice matričnega jezika. Zato se večina tujih glagolov pravzaprav pojavlja v sklopu daljših fraz, ki so v našem korpusu označene kot neuvrščene.

Podatki z zgornjega grafa 5, ki vsebuje zbrane rezultate za vse jezike, nakazujejo, da lahko kot kodni preklp nastopajo vse besedne vrste in da torej za preklapljanje v slovenščini ne veljajo omejitve glede slovničnih besednih vrst, kot jih je predvidel npr. Joshi (1982). Glede na zastopanost jezikov v korpusu preklpov pa lahko tako splošen zaključek pravzaprav naredimo zgolj za angleščino. Če izluščimo angleške členke (priloga C), se izkaže, da je večina le-teh na različne načine zapisanih besed »yup« (jap, jep, jp, jup – skupaj 27 preklpov) in »nope« (nop, nope, noup – skupaj 14), poleg tega pa se pojavijo le še »nah«, »yes«, »right« ter »the«. Med

vezniki se trikrat pojavi »also«, dvakrat »and«, enkrat »or« in »aka«, med zaimki pa so naslednji: »everyone«, »he«, »she«, »what« in »why«. Na podlagi teh rezultatov lahko sklenemo, da je prevzemanje tovrstnih besed res zelo redko, ni pa nemogoče.

Vendarle pa velja na tem mestu izpostaviti še opažanje na podlagi seznama najpogostejših besed v preklopih (priloga E), opisanega v razdelku 3.3.3.3. Tam se namreč izkaže, da imajo med njimi zelo visoko frekvenco prav zaprte besedne vrste. Jasno je, da so te najpogostejše del »otokov vrinjenega jezika« (oziroma »neuvrščenih« preklopov po naši kategorizaciji besednih vrst), in sicer gre za zaimke (»I«, »you«, »my«, »it«), določne in nedoločne člene (»a«, »the«) ter predloge (»to«, »of«, »for«, »on«). Na podlagi tega lahko oblikujemo dodatno hipotezo, da je kombiniranje vsebinskih in sistemskih morfemov znotraj posameznega konstituenta mogoče le izjemoma (ali sploh ni mogoče). Kot že omenjeno v razdelku 2.4.5, bi morali to preveriti v korpusu, kjer bi bile slovnične značilnosti označene podrobneje oz. z dodatnimi kategorijami.

Kljub relativnemu majhnemu vzorcu srbohrvaških in nemških preklopov smo se odločili, da si nekoliko podrobneje ogledamo še pogostost posameznih besednih vrst glede na jezike.

	<i>angleščina</i>	<i>srbohrvaščina</i>	<i>nemščina</i>
samostalniška zveza	33,7 %	20,6 %	24,4 %
neuvrščeno	26,0 %	34,9 %	8,9 %
medmet	16,5 %	1,6 %	0,0 %
prislovna zveza	9,0 %	6,3 %	20,0 %
glagolska zveza	4,1 %	6,3 %	35,6 %
členek	3,6 %	23,8 %	0,0 %
pridevniška zveza	2,7 %	1,6 %	4,4 %
predložna zveza	2,7 %	3,2 %	6,7 %
okrajšava	0,8 %	0,0 %	0,0 %
veznik	0,6 %	0,0 %	0,0 %
zaimek	0,4 %	1,6 %	0,0 %
skupaj	100 %	100 %	100 %

Tabela 11: Deleži preklopov glede na besedno vrsto za tri najpogostejše jezike.

V tabeli 11 lahko opazimo nekoliko višji delež neuvrščenih, daljših besednih zvez in stavkov pri srbohrvaščini. Izkaže se, da je to povezano z večjim številom štiribesednih preklopov, omenjenih v razdelku 3.3.1.3 in vključenih v prilogo B. Podobno izstopa delež členkov – ugotovimo lahko, da je to posledica uporabe dveh besed, saj se med njimi kar 9-krat ponovi »kao«, 5-krat »valjda/valda«, enkrat pa »jeste«.

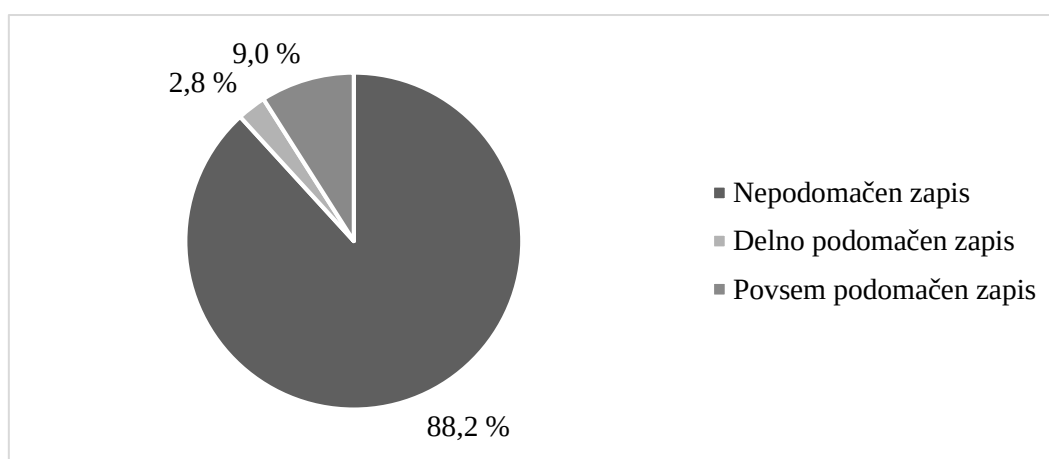
Pri nemških preklopih je zanimivo, da ni niti enega medmeta, precej višja pa sta deleža pridevniških zvez (20 %) in glagolov (35,6 %). Zaradi majhnega vzorca sicer ne moremo

sklepati o splošnih značilnostih preklapljanja v nemščino, smo pa te glagole in pridevnike vključili v prilogo C, saj so zanimivi tudi zaradi svoje izrazite integracije v slovenski (matrični) jezik.

3.3.3 Integracija preklpov v slovenščino

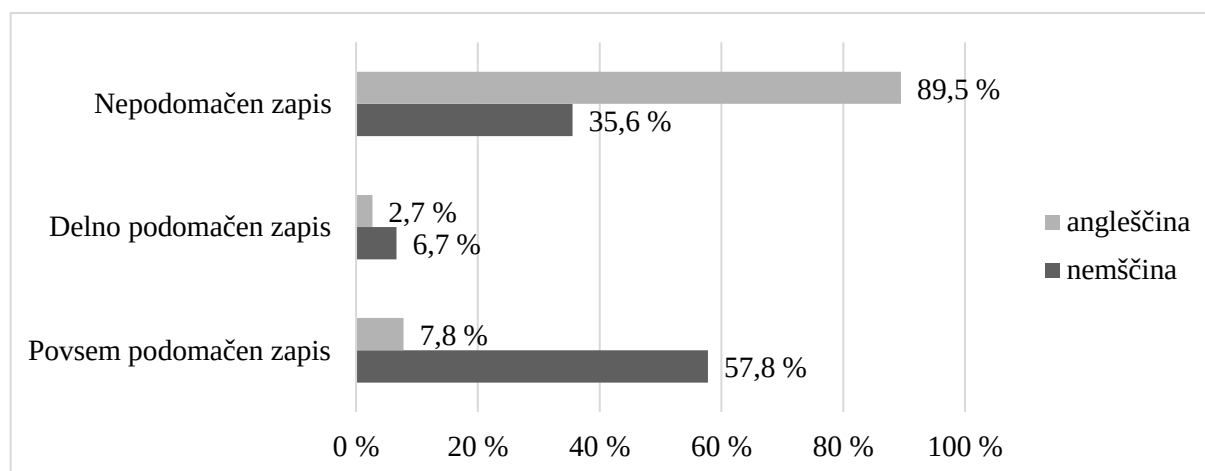
3.3.3.1 *Ortografska integracija (zapis)*

Po preštetju oznak za podomačenost zapisa smo ugotovili, da občutno prevladujejo nepodomačeni kodni preklopi (88 %), torej so besede prevzete citatno, v skladu z njihovim zapisom v izvirnem jeziku (primer: »does a little happy dance«). Kot je razvidno z grafa 6, je bilo delno podomačenih (primer: »velkom back«) in povsem podomačenih (primer: »šejkam«) relativno malo, kar bi lahko govorilo v prid teorijam, da je prav integriranost besede v matrični jezik kriterij za ločevanje prevzetih besed od kodnih preklpov. Vendar je treba izpostaviti, da smo v tem koraku opazovali samo jedro besede, ne pa tudi morebitnih končnic in obrazil. Poleg tega je v literaturi izpostavljena morfofonemska prilagojenost – fonemskega dela pa v našem korpusu ni mogoče preverjati.



Graf 6: Deleži preklpov glede na ortografsko integracijo.

Nadalje smo se odločili natančneje primerjati ortografsko integracijo samo preklpov v angleščini in nemščini, kjer se je izkazalo, da je mogoče večjo stopnjo podomačenosti opaziti v nemških preklpih (gl. graf 7). Predpostavljamo, da je razlog v tem, da gre za besede, ki so jih govorci prevzeli v slovenščino dlje časa nazaj in gre mogoče po drugih kriterijih že za prevzete besede, čeprav jih ni v preskriptivnih priročnikih. Za potrditev te hipoteze se bomo temu vprašanju podrobneje posvetili še v razdelku 3.3.3.2 pri morfološki integraciji nemških preklpov.



Graf 7: Deleži angleških in nemških preklapov glede na ortografsko integracijo.

Po preštetju oznak za podomačenost glede na besedne vrste smo ugotovili, da so najbolj podomačeni členki, ki jim sledijo glagoli, pridevniki, prislovi, medmeti in samostalniki (odstotki so razvidni iz tabele 12). Večjo podomačenost smo pričakovali pri besedah, ki se pregibajo, saj smo predpostavljali, da si bodo uporabniki na ta način poenostavili tvorjenje besedila, a glede na nizek delež podomačenih zapisov lahko ugotovimo, da – predvsem pri angleščini – tviterari pogosto prevzemajo besede v skladu z angleškim zapisom.

	<i>Nepodomačen zapis</i>	<i>Delno podomačen zapis</i>	<i>Povsem podomačen zapis</i>
členek	43,5 %	16,1 %	40,3 %
glagolska zveza	65,8 %	4,1 %	30,1 %
pridevniška zveza	77,9 %	0,8 %	21,4 %
prislovna zveza	87,8 %	0,0 %	12,2 %
medmet	88,6 %	3,2 %	8,2 %
samostalniška zveza	91,8 %	2,5 %	5,7 %
predložna zveza	90,7 %	4,7 %	4,7 %
neuvrščeno	97,8 %	1,3 %	0,8 %
okrajšava	100,0 %	0,0 %	0,0 %
veznik	100,0 %	0,0 %	0,0 %
zaimek	100,0 %	0,0 %	0,0 %

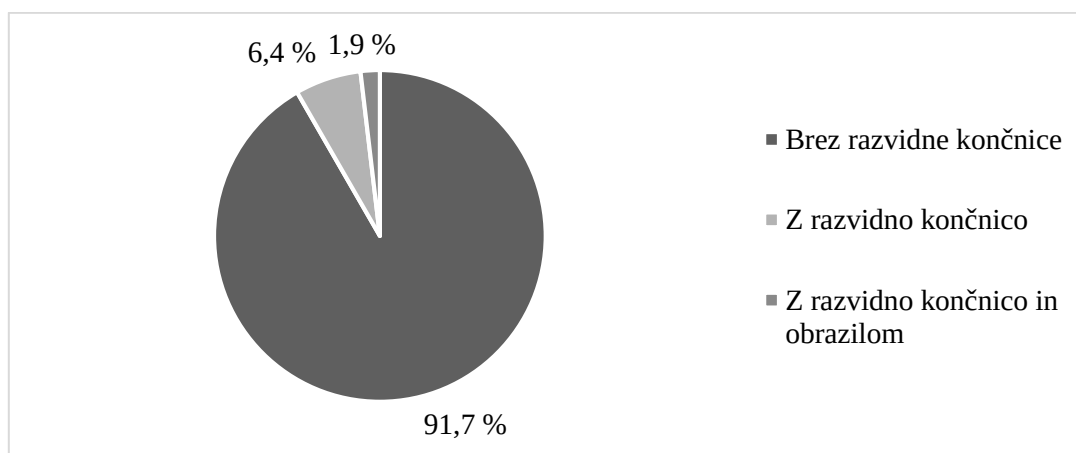
Tabela 12: Ortografska integracija glede na besedne vrste.

Za boljši uvid smo izluščili vse povsem podomačene preklape in jih razvrstili po besedni vrsti in jeziku (navedeni so v prilogi D). Hitro smo ugotovili, da so za velik delež podomačenih členkov »krivi« različni zapisi besede »yup«. Bolj zanimiva kategorija so medmeti, kjer po številu izstopata »sori/sorči« (5x) in »tenks« (3x), med njimi pa so npr. tudi »dbest«, »fakof« in »džiizs«. Med pridevniki po ponovitvah izstopajo besede »kjut« (7x), hepi (3x) in »fakin(g)« (3x), med samostalniki pa »komp«/ »kompjutr«.

3.3.3.2 Morfološka integracija

Na tem mestu naj najprej omenimo, da je v Smernicah za označevanje (priloga A) poleg oznak NR (brez razvidne končnice, primer »hijab«), RK (z razvidno končnico, primer »el clasica«) in ROK (razvidna končnica in obrazilo, primer »unfriendinga«), predvidena tudi oznaka RO (z razvidno končnico). Med označevanjem korpusa smo ugotovili, da sta v njem le dva primera, ki bi ju potencialno lahko označili z RO (»laufer« in »marketingar«), ker pa imata ničto končnico, smo ju označili z ROK.

Po preštetju preklpov brez razvidne končnice oziroma z razvidno končnico in/ali obrazilom smo ugotovili, da je delež neintegriranih preklpov še višji kot glede na zapis (skoraj 92 %; gl. graf 8). Na visok delež zagotovo vplivajo (glede na besedno vrsto) »neuvrščeni« preklpi, ki so tvorjeni v skladu s slovnico vrinjenega jezika, in nepregibne besedne vrste. Poleg tega zagotovo ni zanemarljiv delež samostalnikov z ničto končnico – tovrstni primeri so bili namreč razen zgoraj omenjenih dveh besed, ki pa ne vplivata bistveno na izračunane deleže, označeni kot »brez razvidne končnice«. Odprto tako ostaja vprašanje, ali bi uporabniki te besede v drugačnem kontekstu pregibali v skladu s slovensko slovnico ali pa obstaja večja verjetnost, da se preklpi pojavijo prav na stavčnih mestih, kjer pregibanje ni potrebno.



Graf 8: Deleži preklpov glede na morfološko integracijo.

Pri primerjavi morfološke integriranosti glede na besedne vrste smo upoštevali samo glagole, pridevnike in samostalnike, pri čemer se je po pričakovanjih izkazalo, da so glagoli najbolj podvrženi dodajanju slovenskih končnic in obrazil (kot je razvidno iz tabele 13, ima 56 % glagolov razvidno končnico). Na manjši delež pri samostalnikih vplivajo že omenjene ničte končnice, rezultate za pridevniške zveze pa je mogoče pripisati dejstvu, da so številni pridevniki pravzaprav del samostalniških fraz oziroma t.i. »otokov vrinjenega jezika«. Morda obstajajo

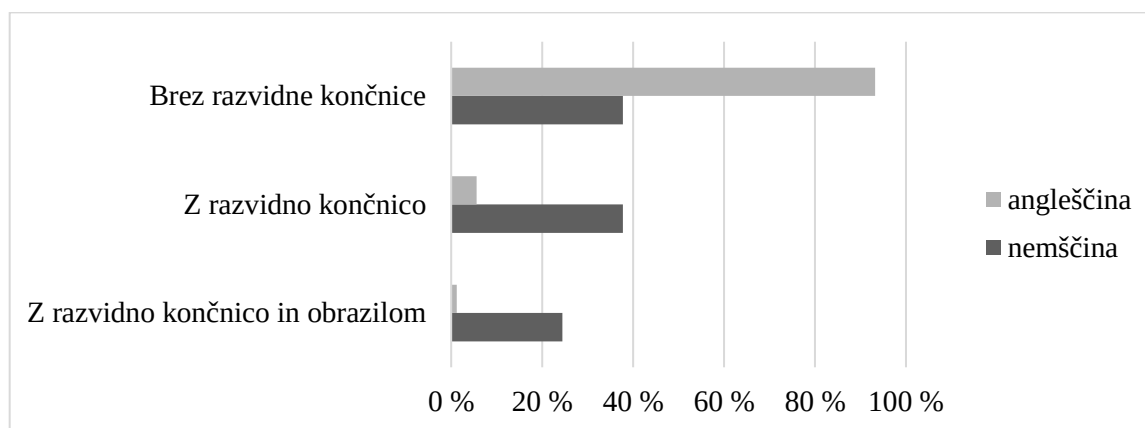
določene slovnične omejitve pri preklapljanju pridevnikov na mestih, kjer bi v skladu s slovensko slovnico morali dobiti ustrezno končnico. Vendar pa v naši raziskavi nismo označevali besednih vrst za vse besede, zato te hipoteze ne moremo empirično preveriti.

	<i>Brez razvidne končnice</i>	<i>Z razvidno končnico</i>	<i>Z razvidno končnico in obrazilom</i>
glagolska zveza	44 %	45 %	11 %
pridevniška zveza	94 %	0 %	6 %
samostalniška zveza	85 %	13 %	2 %

Tabela 13: Morfološka integracija glede na besedne vrste.

Podobno kot zgoraj smo tudi tu izdelali preglednico z morfološko integriranimi preklopi, razvrščenimi po besedni vrsti in jeziku (vključena v prilogi D). Po številu ponovitev med glagoli izstopajo »laufati«, »šerati«, »filati«, »kenslati«, »lajkati«, »potegati«, »rentati«, med samostalniki pa se dvakrat ponovijo le »follower«, »komp«, »taxi« (v različnih oblikah). Na splošno lahko ugotovimo, da pride do morfološke integracije pri pestrejšem naboru besed kot pri ortografski ter da tviteraši nimajo posebnih težav z dodajanjem slovenskih končnic tudi na nepodomačene osnove – le v petih primerih so za to na primer uporabili vezaj.

Po analogiji z argumenti pri analizi ortografske integracije smo tudi pri morfološki podrobneje primerjali samo angleščino in nemščino. Tudi na grafu 9 je očitna razlika med razporeditvijo deležev, in sicer je v nemščini precej več preklpov z razvidno končnico in/ali obrazilom, kar nadalje podpira našo hipotezo, da gre za starejše, bolj integrirane preklope, ki so morda na poti, da postanejo prevzete besede (ali pa so to že glede na rabo).



Graf 9: Deleži angleških in nemških preklpov glede na morfološko integracijo.

Na tem mestu lahko izpostavimo tudi problematičnost našega izbranega kriterija za ločevanje prevzetih besed od kodnih preklpov. Med označevanjem se je izkazalo, da so številni

germanizmi morda proti pričakovanjem že vključeni v priročnike, npr. cajt, frišne (roke) v SSKJ, pucati, fuzbal, luft, šporhet v Sloleksu ali šmirglati v SP, zato smo jih kljub morebitnim kvalifikatorjem (npr. nižje pogovorno) šteli kot del slovenskega besedišča in jih nismo označevali kot kodne preklope. Razlog za njihovo večjo integracijo se zagotovo skriva v tem, da je zgodovina nemško-slovenskih jezikovnih stikov daljša v primerjavi z angleščino, saj sega vse do druge polovice 8. stoletja, nemški kulturni vpliv na slovenščino pa je začel pojenjati po letu 1917 (Šekli 2015: 41–42).

3.3.3.3 *Ortografska in morfološka integracija*

Če iz korpusa preklonov izluščimo preklope, ki imajo hkrati oznako za (delno) podomačenost zapisa ter razvidno končnico in/ali obrazilo, dobimo vzorec 50 preklonov, predstavljenih v prilogi D, med katerimi je šest pridevniških (npr. »pofarban«), 24 glagolskih (npr. »hajcajo«, »potegal«) in 20 samostalnih zvez (npr. »frendi«, »kompu«). V celotnem korpusu preklonov to predstavlja 3,47 % vseh preklonov. Po kriteriju integracije bi lahko bile te besede kandidatke za »prevzete besede« oziroma prevzem v slovarje, pri čemer bi seveda morali dodatno upoštevati tudi pogostost njihove rabe, kar pa bi morali preverjati v večjem korpusu, kot je naš. Lahko pa so ti rezultati izhodišče za kakšno drugo raziskavo v tej smeri.

Kljub temu smo se vprašanja, kateri preklopi bi morda še lahko bili potencialne prevzete besede, dotaknili še z druge plati, in sicer tako, da smo izdelali seznam najpogostejših besed iz vseh preklonov (ne glede na to, ali se pojavijo samostojno ali kot del daljše zveze), ki je vključen v prilogi E. Vključuje besede, ki so se pojavile vsaj trikrat, pri čemer so upoštevani različni zapisi (npr. »lol« z različnim številom o-jev, »thanks/tnx/thx«, »happy/hepi«). Če je njihova raba pogosta tudi v drugih situacijah, bodo nekatere izmed njih morda sčasoma morda prešle med prevzete besede.

K celotnemu razdelku 3.3.3 pa moramo dodati opombo, da je na rezultate gotovo vplival tudi način vzorčenja, pri katerem smo zajeli tvite, ki vsebujejo besede z oznako Nj, avtomatski označevalnik pa morda tuje besede s podomačenim zapisom zazna kot slovenske in jih ne označi z Nj.

3.3.4 Razlike glede na spol

V označenem korpusu je 730 preklonov v tvitih ženskih avtoric (53 %) in 651 moških (47 %), vendar glede na majhno razliko in velikost vzorca ne moremo trditi, da ženske preklaplajo pogostejše kot moški. Ob tem moramo izpostaviti, da v celotnem podkorpusu twitov Janes prevladujejo moški (53 % v primerjavi s 24 % žensk, ostali pa so označeni kot nevtralni).

Pri primerjanju zastopanosti jezikov glede na spol smo ugotovili, da je pri moških malenkost višji delež srbohrvaških preklpov (5,4 % v primerjavi s 3,8 % pri ženskah), a tudi tu ne gre za signifikantno razliko. Glede dolžine se izkaže, da so moški tviti v povprečju nekoliko daljši (2,4 besede pri moških in 2,14 pri ženskah), čeprav imajo nekoliko večji delež znotrajstavčnih preklpov, ki so načeloma krajši (71 % znotrajstavčnih pri moških, 63 % pri ženskah). Podrobnejši podatki so v tabelah 13, 14 in 15. Zaključimo lahko, da z obzirom na kodno preklapljanje med obema spoloma ni večjih razlik pri komunikaciji na Twitterju.

	<i>Ženske</i>	<i>Moški</i>	<i>Skupaj</i>
Število (delež) preklpov	730 (53 %)	651 (47 %)	1381
Število (delež) medstavčnih preklpov	267 (37 %)	191 (29 %)	458 (33 %)
Število (delež) znotrajstavčnih preklpov	463 (63 %)	460 (71 %)	926 (67 %)

Tabela 14: Pogostost preklpov v tvitih ženskih in moških avtorjev.

	<i>Ženske</i>	<i>Moški</i>	<i>Skupaj</i>
angleščina	91,0 %	89,1 %	90,1 %
srbohrvaščina	3,8 %	5,4 %	4,6 %
nemščina	3,2 %	3,4 %	3,3 %
latinščina	0,4 %	0,9 %	0,7 %
španščina	0,7 %	0,6 %	0,7 %
francoščina	0,4 %	0,5 %	0,4 %
italijanščina	0,3 %	0,2 %	0,2 %
arabščina	0,1 %	0 %	0,1 %
poljščina	0,1 %	0 %	0,1 %
skupaj	100 %	100 %	100 %

Tabela 15: Deleži jezikov v tvitih ženskih in moških avtorjev.

	<i>Ženske</i>	<i>Moški</i>	<i>Skupaj</i>
1-besedni	50,3 %	50,8 %	50,5 %
2-besedni	22,5 %	21,0 %	21,8 %
3-besedni	10,4 %	13,8 %	12,0 %
4-besedni	8,5 %	6,1 %	7,4 %
5-besedni	3,3 %	2,9 %	3,1 %
6-besedni	1,6 %	2,5 %	2,0 %
7-besedni	1,9 %	0,8 %	1,4 %
8- do 13 besedni	1,5 %	2,1 %	1,8 %
skupaj	100 %	100 %	100 %

Tabela 16: Dolžina preklpov v tvitih ženskih in moških avtorjev.

3.4 Kvalitativna analiza

Na podlagi klasifikacij, predstavljenih v teoretičnem uvodu, smo sestavili nabor kategorij, ki smiselno ustreza opažanjem, zbranih med označevanjem in dodatno analizo kodnih preklapov. V nadaljevanju predstavljamo nekaj ilustrativnih primerov kodnih preklapov za vsako kategorijo skupaj s kontekstom, v katerem se pojavijo. Seveda bi bilo mogoče posamezne tvite uvrstiti tudi v več kategorij, saj se lahko funkcije in teme prekrivajo ali pa ker vsebujejo več različnih preklapov. V zvezi s slednjim smo se tudi odločili, da v vsakem primeru podčrtamo zgolj tisti preklap, ki je relevanten za vsako kategorijo.

Primere smo dodatno anonimizirali tako, da smo izbrisali oznake z imeni uporabnikov (@...), kadar to ne vpliva na skladenjsko celovitost tvita, v nasprotnem primeru pa smo uporabnika anonimizirali v @tviteraš, zavaljo večje preglednosti pa smo spletne povezave okrajšali na http.

3.4.1 Diskurzne funkcije kodnega preklapljanja

3.4.1.1 *Zgoščeno izražanje (gospodarna raba jezika)*

Na Twitterju je zaradi prostorske omejenosti gospodarna raba jezika še pomembnejša kot sicer, zato ne preseneča, da si uporabniki pri tem včasih pomagajo tudi s tujimi jeziki. Angleščina je v tem pogledu zelo priročna, saj omogoča nizanje samostalnikov v obliki levih prilastkov, kar načeloma v slovenščini ni mogoče. Primeri 1–5 prikazujejo, kako so se avtorji elegantno izognili daljšim pojasnjevalnim odvisnim stavkom.

- [1] Optimalno instagramabilna garderoba za vse ljubitelje teh srečnih "do more of what makes you happy" citatov 😊 http
- [2] Če damo v enačbo še "priročnost" in "stroškovno učinkovitost", digital čisto povozi paper-based content rezultate. http
- [3] ehm. cough, cough.. gag... jojoj, neki se mi je zataknilo... valda ne bom zdaj še old ppl kašlja dobila?!?!?
- [4] Pravkar sem odkril dismiss gumb. *does a little happy dance*
- [5] To je state capture stanje, ki z etiko/moralo prebivalstva nima veze. N.Plavšak je spet v sedlu, zadeva se nadaljuje

3.4.1.2 *Ponovitve*

Kljub kratkosti besedil na Twitterju pa lahko avtorji uporabijo dva izraza v različnih jezikih za isto stvar. Motivacije za to so lahko različne, npr. dodatna pojasnjevalna vrednost tujega termina [6], kulturna specifičnost pojma, ki potrebuje še slovensko razlago [7], nagovarjanje različnih občinstev [8] ali pa morda na ta način podkrepijo svoje sporočilo.

- [6] Swine, swine ... morda tudi ptičja (avian flu) - skratka eno super zdravilo:) http
- [7] Ti, kok se že imenuje tist čaj? "Barut, se mi zdi," odgovori misleč na bure baruta, oziroma sod smodnika. #gunpowder
- [8] Nocoj v rojstni hiši Huga Wolfa v Slovenj Gradcu / Hoy en la casa natal de Hugo Wolf en Slovenj Gradec, Eslovenia http

[9] Ma kakšnega proška :-P. Prosecco! Suh.

[10] Dnz spim s slovarjem, da bova jutri na ti. :D Lahko noč oz. buenas noches! :) #španščina #matura

3.4.1.3 Odzivanje oz. navezovanje

Uporabniki včasih posebej zaznamujejo tujejezični del besedila z navednicami, kar kaže, da ga občutijo kot nek tujek oziroma ga ne vključijo nezaznamovano v svoj diskurz. Razlog za to je lahko, da se navezujejo na predhodnika [11, 15] ali se odzivajo na druge zunanje okoliščine [12, 13, 14]. Kljub temu da gre za dobesedne navedke iz drugega konteksta, smo takšne primere označili na enak način kot ostale preklope, saj ustrezajo izhodiščni opredelitvi (kodno preklapljanje je uporaba več jezikov v istem stavku/diskurzu).

[11] Pa ja. Sej če odzameš "Praise the Lord!" del, je čist OK :) #keepemcomin

[12] sovražim reklame z namenom približat se mladini z uporabo slenga pa besed kot npr. "kul" "yo" "ful" "dbest" "keš" #fuj

[13] Ne, "for your information" ne morete kar dobesedno prevajati v "za vašo informacijo". #sampravm

[14] Izraza "YOLO moment" še nisem slišala, mi je pa takoj prirastel k scrui :)

[15] al pa "ooooor, you know" in naprej po slovensko :P

3.4.1.4 (Retorična) vprašanja

Med daljšimi, predvsem medstavčnimi preklopi smo opazili kar nekaj vprašanj. Tudi tukaj je lahko razlog težnja po gospodarnejši rabi jezika ali pa želja po večjem poudarku, morda pa želijo uporabniki na ta način vzbuditi dober vtis pri ciljnih bralcih oziroma je tovrstno preklapljanje del njihovega osebnega sloga. V zvezi s tem bi zanimiva tudi kvalitativna analiza večjega števila tvitov istega avtorja, s katero bi lahko ugotovili, ali oziroma kako se slog preklapljanja razlikuje med različnimi uporabniki.

[16] Skor bi leče dala v micelarno čez noč. How am I a grownup?

[17] Tečna. Grem pečit. Muffins, anyone?

[18] what's there to tell? Zvečer greš v savno, si tm do konca, pol pa domov:D

[19] Jaz se še kar držim starejše TOSHIBE z Visto in ne menjam, dokler ne crkne žvauca. Why fix/change if it ain't broken? :)

[20] si ti prebral lasten link in če si, what's your point? Ali pa si prebral samo naslov?

3.4.1.5 Odgovor sogovorniku

Že iz seznama najpogostejših besed v prilogi E je bilo razvidno, da se v preklopih pogosto pojavlja beseda »you«, poleg nje pa so tudi primeri z nestandardnim zapisom tega zaimka (»u«), tako da je preklopov, ki potencialno spadajo v to kategorijo, več kot 40. Na Twitterju se lahko odvijajo pravi pogovori, zato bi bilo za natančnejšo analizo, kakšno vlogo igrajo preklopi v teh primerih, morali upoštevati še predhodne tvite, kar pa v okviru naše raziskave zaradi načina vzorčenja ni mogoče. Tako na tem mestu samo prikazujemo le nekaj zanimivih primerov, ki kažejo, kako se v pogovorih med tviteraši prepletata slovenščina in angleščina.

[21] Upočasnjjen promet vlakov zaradi vročine. Ker se tiri zvijajo. Now you know.

- [22] Yes you are, grumpy. Sem ti odgovorila. Ti se pa zjasni, a iščeš fotko, ali izraz
 [23] I see your nič nimam za oblečt and raise you nič nimam za obut. #endofseasonproblems
 [24] pejt v dnevno, daj zvočnike do konca na glas ... in pol pejt vn. Can't do it, can you? :P pridem popoldne :D
 [25] u and me both...danes pogoltnem uspavalo pa bo :D kdaj se vračaš?

3.4.1.6 Vljudnostni izrazi

Tudi za to kategorijo najdemo utemeljitev na seznamu najpogostejših besed (Priloga E). Predvsem pri različnih okrajšanih zapisih besede »thanks« lahko sklepamo, da k izbiri pripomore težnja k čim hitrejšemu in čim krajšemu sporazumevanju, kar je tudi na splošno ena od značilnosti računalniško posredovane komunikacije. »Sori« je prav tako krajše od slovenskega »oprosti«. Poleg tega lahko pogosto pojavljanje teh izrazov povežemo tudi z večjo neformalnostjo komuniciranja na Twitterju.

- [26] please, ne, ne more, niti 1 % šanse. On je v igri zaradi drugih interesov. Tudi R nominacije ne more dobiti, kaj šele volitve.
 [27] lej sorry sej te nisem mislu tolk napadat kot je vceri izpadlo ampak trenutek za tvoj post ni bil na mestu ...
 [28] joj sori :D Ne res je fajn gledati kako v teh časih uspevata le dva biznisa: tisti low budget in pa tisti beyond imagination...
 [29] aha tnx ... dvomim da bo z avtom tko da ni panike ...
 [30] a, thank you, mogoče pridemo kej sparglje degustirat :)

3.4.1.7 Izražanje čustev

Na seznamu najpogostejših besed je prvo mesto zasedel zaimek »I« (jaz). Tviteraši torej takrat, ko preklaplajo v angleščino, pogosto povedo nekaj o sebi ter na ta način tudi izražajo svoja čustva in trenutno razpoloženje, kar je razvidno tudi iz primerov 31–35.

- [31] ze doma??? Oh how i envy you!!!! :p
 [32] Folk hodi na medene tedne. Midva bova kupila parking in kamin za v flat. And I couldnt be happier!
 [33] A veš uno ko si general za situacijo za nazaj in se pol znajdeš v točn taki situaciji, pa ponoviš začetniško napako. I will never learn.
 [34] i love you! Ful ss manj nesposobbo pocutim :)
 [35] Oh kako zelo odraslo od nje, res. Podpihovanje drugih proti meni, kot da smo se v os. Ljubica, I don't care, ti kar. :)

3.4.1.8 Medmeti – klepetalniške krajšave

V kvantitativnem delu raziskave smo že ugotovili, da so med preklopi zelo pogosti angleški medmeti, med katerimi je veliko klepetalniških krajšav, ki so posebej značilne za računalniško posredovano komunikacijo. Na seznamu najpogostejših besed se je tako na vrhu znašla beseda »lol« (angl. *laugh out laughing*), med tistimi z veliko pojavitvami pa so še »omg« (angl. *oh my god*), »btw« (angl. *by the way*) in »wtf« (angl. *what the fuck*). Spodaj so prikazane v kontekstu celega tvita.

- [36] Lol :D Sošolka si je vedno želela 3 otroke: Žak/Pak/Mak. No, Žaka že ima :)

- [37] LOOOOOL a je tudi tebi nek 'Buy Fol-l-owers 1K-\$12' favorital tvit?
 [38] omg. pa kje ti ljudje rastejo? :D
 [39] Na nebotičnik sva odšla... WTF, a Angleže hočemo prestrašiti z baladami? :P
 [40] ma deej. Tam sm ze tut bla btw :p

3.4.1.9 Ekspresivni elementi

Tviteraši pogosto posežejo po angleščini, ko želijo še posebej poudariti svojo izjavo. Morda je preklinjanje v tujem jeziku bolj sprejemljivo za ciljno občinstvo oziroma se zdi tudi uporabnikom manj vulgarno, po drugi strani pa morda slovenščina ne ponuja avtohtonih izrazov, ki bi ustrezno zajeli in prenesli želeno sporočilo.

- [41] če je konc, je faking konc. Nič ni večno. Prosim, pejm napreeej! ;)
 [42] moja zena tud ne, pa je faking hot ;)
 [43] Skenslaš kavo na 2 šalce na dan, pa te glava boli vsak dan za popizdit. Fuck this shit. Kadit je blo loži nehat o.O
 [44] Ji naštimam frišne brikete pa vodo, noče pojest. Zbrska brikete v vodo, ewo njo, poje takoj pa si misli fuck, ker nesposobn staff. #cats
 [45] bemtiš če bi žvel nekje ob meji, madona da bi s principa tankal čez mejo. Fakof, res!
 [46] Holy shit, Ljubljancani. Vse lovorike za najbolj nesposobne voznike poberete vi. Mi je žal.
 [47] prlet v Lj damnit!
 [48] damn right! Sem vedu da ti bo vsec ;)
 [49] Bullshit! Makeup, cunje, razsvetljava, kamera, računalnik. Ozri se okoli sebe, princeska!
 [50] damn internet connection, cant even see my boo :(@tviteraš mogoče je pa čas da se odpravim iskat justina :D

3.4.2 Tematska polja

3.4.2.1 Znani citati

V raziskovalnem korpusu so primeri, kjer tviteraš uporabi frazo v tujem jeziku, ker citira tuje besedilo. Na ta način je citat bolj zaznamovan, kot bi bil v prevodu, morda bi ga bilo težko prevesti ali pa bi se s prevodom pravzaprav izgubila aluzija, ki jo avtor poskuša doseči.

- [51] Zakaj Nemčija(Merkel:Islam ist ein Teil von Deutschland)noče sprejeti tega dela,ki ji trka na vrata?Še več,podkupuje polit.&ga vsiljuje drugim
 [52] Prijaznost pripelje le do neke tocke. Saj ves, good girls go to heaven, bad girls go everywhere else.
 [53] Elementary, Dr. Watson: investicije se ne more krasti kar tako, kredite se raztala prijateljem kot karte.
 [54] Ej,neka se piše, pa makar i dobro! (Bregović)Včasih sem jokala zaradi laži o sebi.Danes sem skala.Vrnem pa, če je treba;)
 [55] V #GOT niso imeli pojma, pravilno bi bilo "Spring is coming" :)

3.4.2.2 Televizijske serije, filmi

Med označevanjem je bilo opazno, da se tviti, ki vsebujejo kodno preklapljanje, velikokrat vrtijo okrog popularnih televizijskih nanizank ali nadaljevanj oziroma filmov. Glede na to, da je angleščina prav prek teh medijev postala tako vplivna tudi v slovenskem prostoru, nas to ni presenetilo. Očitno pa nekatere med pisanjem o tej temi v povezavi z določenim besediščem

kar potegne v angleščino, v določenih primeri pa morda res nimajo na voljo primernih slovenskih ustreznice – vsaj ne tako kratkih in uveljavljenih oz. razumljivih (primera 58 in 59).

[56] dolgočasn psychological horror. Tečn mulc za popizdit!! 🐍 @tviteraš mi je biu všeč. A breath of fresh air :)

[57] ane, čist underrated serija. poglej še Uk verzijo... Sam men niso bli ušeč :)

[58] no sej, spet ne vem ker dan je, navadn. sobota je drgač že. 😊 jst pa gledam zakaj ni streama nikjer, lol.

[59] jup. + Alison pa Felix rabta spinoff 😊

[60] mislm da smo tle kr vse weirdos. Dejmo se kregat okol fictional characters no :D

3.4.2.3 IT, internet

Slovenščina se prilagaja razvoju novih pojavov in tudi za spodaj prikazane primere bi zagotovo našli povsem spodobne slovenske ustreznice. Kljub temu ne preseneča, da je kar nekaj kodnih prekopov povezanih z novimi tehnologijami in internetom. Osebe, ki se denimo ukvarjajo s spletnim marketingom, o tem zagotovo veliko berejo v angleščini, prav tako verjetno uporabljajo programe in spletna orodja, ki niso lokalizirana v slovenščino, zato so jim angleški izrazi bolj domači. Nekateri pa imajo pred slovenskimi prednost tudi v tem, da so precej krajši, npr. »komp« za računalnik ali »app« za aplikacijo.

[61] Si poskusil ne-obnovit backupa morda? Ali takrat deluje? Ko je cisto sveza masina?

[62] Kako napisati scam e-mail? Slovensko - Nigerijski poslovni forum.

[63] Hmm gledam ja, samo če idejo pustimo na stran...kolko efekta pa to ima? Če je viral success nekaj tisoč views..moh.

[64] nisem mailchimp a nikol uporabljal. Ze tisti cartoon vmesnik me odbija :) a to za autoresponder rabis ali classic newsletter?

[65] resno, web player zalaufi, ne app na kompu.

3.4.2.4 Šport

Šport je gotovo ena bolj priljubljenih tem, zato ne preseneča, da smo jo uvrstili tudi v naš izbor kategorij. Do kodnega preklapljanja na tem področju pride predvsem zaradi navijanja za tuje klube, kot prekop pa smo šteli tudi veznik »versus« (zapisan kot vs ali v), ki se pojavlja npr. pri navajanju tekem iz angleške nogometne ali ameriške košarkarske lige.

[66] Fino, Arsenal. Zdaj v pričakovanju el clasica. #nervous

[67] Marguč v igri za novica sezone v LP. Glasujmo, da postane Rookie of the season! :) http

[68] Ob navijanju komentatorjev za odpoved veleslaloma in druge serije poletov vsaj naši športniki dajejo vzgled. Bravo #orli #fairplay.

[69] V soboto se oglasim s tekme Crystal Palace vs. Chelsea #CPLCHE #SelhurstPark Lani je ta obračun dobil Palace!

[70] Jutri bo zmagala, ker bo jezna! :) GO @TinaMaze! #Are

3.4.2.5 Kulturno specifični pojmi

Različni kulturno pogojeni pojmi so pogosto trd oreh še za prevajalce, tako da ne preseneča, da je s tega področja precej kalkov ali citatno prevzetih besed. Med takimi, ki se v slovenščini še

niso uveljavile do te mere, da bi bile vključene v priročnike, so tudi spodnji preklopi, med katerimi prevladujejo razne jedi.

- [71] Od 09:30 v kuhinji...Cheesecake v hladilniku, predjed končana, ostalo pa pripravljeno, da skoci v lonce....jupi;)
- [72] Kdo ve kje kupiti 'plantain' banane? Ki niso banane... <http>
- [73] lahko probaš tud sam nardit :) Horaa Osba'o
- [74] uf, good luck. jst sm red velvet kapkejke jedla. biskvit, pofarban na rdeč. ne štekam hajpa razn tega da je pretty.
- [75] Čemu je predsednica srbskega parlamenta na nedavnem obisku v Iranu nosila hijab? Jo je morda zeblo?

3.4.2.6 Twitter

Izmed vseh tematskih polj smo za tega našli največ primerov, predvsem zaradi razširjene uporabe kratice RT, ki smo jo že izpostavili v teoretičnem delu, pojavita pa se tudi MT (angl. *modified tweet*) in FT (angl. *follow-up tweet*). Pogoste so besede z jedrom *follow* (slediti) in *handle* (angleški izraz za oznako, sestavljeno iz znaka @ in imena naslovnika). Pod to kategorijo pa uvrščamo tudi tvite, v katerih se pojavi *selfie*. Vsekakor pa ni presenetljivo, da na Twitterju govorijo tudi o Twitterju.

- [76] A notification si dobil za to? Al so to kakšni plačani followi, da se ti ne prikaže med notificationi?
- [77] Hmm, če kdo pogleda tvoj tw account, bi se mu ziher zazdelo, da je imenovati te Haso še kar blaga oznaka za tako sarma yugoljublje.
- [78] tvoj handle ti pristoji 😊
- [79] Mi smo old-farts, ki smo se navadili na RT, še preden je TW vpeljal to "retweet" novotarijo :-)
- [80] O vau, skoraj 200 followerjev. Vedno več ljudi ne motijo moji bulšit tviti. Tenks. :*

3.4.3 Frazeologija

Preučevanje različnih frazeoloških enot, ki se na Twitterju pojavljajo kot kodni preklopi, bi si zaslužilo samostojno raziskavo. Na tem mestu tako predstavljamo samo nekaj zanimivih primerov, ki so pritegnili našo pozornost med označevanjem drugih značilnosti. Ti odstirajo iznajdljivost in jezikovno znanje, ki ga tviteraši s pridom uporabijo za prenašanje zelenega sporočila. Še posebej so pogosti različni idiomi – »makes sense« se v našem korpusu npr. pojavi štirikrat pri štirih različnih avtorjih. Razlog za to je gotovo ta, da je idiomatske besedne zveze velikokrat težko ustrezno prevesti v drug jezik, njihova pogosta raba med tviteraši pa kaže na njihovo dobro znanje angleščine. Poleg tega so frazemi del mentalnega leksikona in jih govorci priključijo iz spomina kot celoto (Jakop 2006: 31).

3.4.3.1 Pregovori

- [81] zakon! :) (pa skromno pripominjam: greqt minds think alike :))
- [82] Torej lahko sedaj spet napišem za Majdo Potrata: Nomen est omen. Spet Potrata davkoplačevalskega denarja.
- [83] u snooze u lose. Sem ful tulila, kdo bi.. udej je dobu

[84] majo pravico, to the victor go the spoils, škodo pa samo sebi delajo, naše civilizacije ne morejo tko niti mal prizadet, obratno

[85] Če ne še dlje. Sam ne zdržim jih dolg nosit. Sej več, less is more :)

3.4.3.2 Idiomi

[86] Brad Pitt na Planet TV. U are welcome ;)

[87] I'm not feeling myself today. Čudne izjave letijo iz mojih ust. Mogoče rabim Snickers. Mogoče...

[88] fresh meat. Luščna, dober okus! Pa brihtna tudi ni lih, a? ;)

[89] a si se odločil, da boš couch potato ratal? :)

[90] Dajmo definirat: trema pride, ko ti je nekaj še posebno pomembno in bi se rad še posebno izkazal. Give them a break. Saj bo :) #junaki

[91] there you go :) sicer pa nikoli ne veš, mogoče so bili kandidati, ampak ne ustrezajo strukturi

[92] Eh, pogojno tri leta. A slap on the wrist... RT @Delo: Gordani Kalan Živčec 10 mesecev zapora http

[93] On a side note: Benti, nove baterije za UPS rabim :) #prejsnjitvit

[94] Highfive, danes si čist on fire! Zapisujem plus pike! :D

[95] dolgočasn psychological horror. Tečn mulc za popizdit!! 🐞 @tviteraš mi je biu všeč. A breath of fresh air :)

3.4.3.3 Frazni glagoli

[96] Še malo pa bo ponedeljek..., so bring it on!!! http

[97] Toooo!!! A čutiš to, a? To je ta lahkota inbox zero! Keep it up!

[98] šment, sem imela v mislih stajling z balerinkami in enim retro kabrioletom :) oh well, dream on

[99] It goes on. Ustvaril jo je arhitekt Jeza, asistent pa je bil - Triler.

[100] Eh, chickened out eventually. Ugotovil, da imam samo neizolirano olfo. Pa sem šel raje varovalko ven odšraufat. Pussy. #prejsnjitvit

3.4.3.4 Stalne besedne zveze in kolokacije

[101] my bad, na drugo sem pozabila :(

[102] če boš kdaj usput mim švedske hodila pridi na makeup :) free of charge. Sam prnes uno hrustljavo pito :P

[103] moja je šla for good, sam še na obiske bova hodile :S nevem kwa je hujš, k gre desc al frendica.

[104] 15min. Več k to je waste of (my) time.

[105] hihi, my thoughts exactly. Ceprav razmislijam, da tko okorno hodi zato, da jo lazje fotkne. Recimo :D

4 Zaključek

V raziskavi smo ugotovili, da kodno preklapljanje v tvitih slovenskih uporabnikov vsekakor ni redek pojav, čeprav samo na podlagi naših ugotovitev ne moremo oceniti absolutne pogostosti preklapljanja v slovenskih tvitih na splošno. Predvidevamo lahko, da do podobnega mešanja jezikov prihaja tudi v drugih oblikah nestandardne (spletne) komunikacije, ki bi jih morali preučiti na podoben način kot v naši analizi, preden bi lahko z večjo gotovostjo rekli, kakšne so zakonitosti kodnega preklapljanja v slovenščini na splošno. Kljub temu lahko že na podlagi predstavljene analize besedil s Twitterja ugotovimo, da gre za zelo raznolik pojav.

V našem korpusu so po pričakovanjih prevladovali znotrajstavčni preklopi (67 %), kljub prostorski omejenosti besedil pa tviti v povprečju vsebujejo 1,26 preklopa, torej lahko uporabniki preklopijo iz tujega jezika nazaj v slovenščino, preden znova posežejo po tujejezični prvini. V tvitih prevladujejo enobesedni preklopi (približno polovica vseh preklapov), povprečna dolžina preklapov pa je 2,13 besed (1,73 v znotrajstavčnih in 2,94 v medstavčnih). To je v skladu z našo hipotezo in ugotovitvami iz literature – enobesedne preklape namreč zlahka uporabijo tudi osebe, ki morda niso tako vešče tujega jezika, medtem ko je za prenos daljših, celo frazeoloških besednih zvez potrebnega že nekaj več jezikovnega znanja. Poleg tega je med preklopi veliko t.i. klepetalniških krajšav, značilnih za spletno komunikacijo (npr. »lol« ali »omg«), ki imajo vlogo medmeta ter izražajo avtorjevo razpoloženje in odnos. Potrdili smo tudi hipotezo, da je v nestandardnih besedilih delež preklapljanja višji.

Med jeziki je izrazita dominantnost angleščine (90 %), ki ji sledita srbohrvaščina (4,6 %) in nemščina (3,3 %). Preklapov iz ostalih jezikov je zelo malo, vendar smo zaznali 9 različnih jezikov. Pri vsebinski analizi teh tvitov se je izkazalo, da gre pri drugih jezikih predvsem za formulaične izraze in kulturno specifične pojme, angleščino pa uporabniki uporabljajo na bolj raznovrstne, tudi kreativne načine. Tako izrazit delež angleščine je mogoče pripisati več razlogom – predvsem mlajše generacije jo praviloma dobro poznajo že iz osnovne in srednje šole, vseprisotna je v popularni kulturi (filmi, glasba), dodatno pa širi svoj vpliv tudi kot glavni jezik interneta.

V zvezi s tem bi bila zanimiva primerjava, kakšna je zastopanost jezikov v drugih (spletnih) medijih, denimo na forumih, blogih ali v spletnih komentarjih, saj bi na ta način lahko ugotovili, ali je tako pogosto preklapljanje iz slovenščine v angleščino značilno za računalniško posredovano komunikacijo na splošno ali gre za specifično lastnost komuniciranja na Twitterju (kot zanimivost: 9,6 % tvitov v podkorpusu tvitov Janes je bilo identificiranih kot v celoti

objavljenih v angleškem jeziku, kar je največji delež tujejezičnih vsebin v celotnem korpusu, saj ostali viri prispevajo bistveno manj: Wikipedija 2 %, ostali pa pod 1 %). Z bolj sociolingvističnim pristopom pa bi lahko preverili, kakšen je odnos avtorjev in bralcev do kodnega preklapljanja. Predvsem bi bilo zanimivo videti, ali se odnos spremeni glede na jezik preklopa ter kakšne so morebitne medgeneracijske razlike pri sprejemanju tujejezičnih prvin.

Hipotezo glede zastopanosti besednih vrst smo potrdili le deloma; med preklopi je bil najvišji delež samostalniških besednih zvez (33 %), drugo mesto pa so zasedli medmeti (15 %), predvsem zaradi že omenjenih klepetalniških krajšav. Deleža pridevniških besednih zvez (9 %) in glagolov (5 %) sta bila nekoliko nižja od pričakovanj, na kar je vplival precejšen delež »neuvrščenih« preklapov (26 %), tj. daljših zvez, v katerih so glagoli pogosto del stavka, tvorjenega v skladu s slovnico tujega jezika, poleg tega so pridevniki lahko tudi del samostalniških besednih zvez.

Zbrani podatki ne zadostujejo za sklepanje o tem, ali za kodno preklapljanje v slovenščini veljajo katere od slovničnih omejitev, ki so jih raziskovalci odkrili v drugih jezikih. Za jezikovni par slovenščina-angleščina smo ugotovili, da se lahko kot kodni preklop pojavijo tudi angleški členki, vezniki in zaimki, čeprav bolj poredko. Predpostavili bi torej lahko, da za to kombinacijo prej velja načelo »vse je mogoče« oziroma da lahko uporabniki s svojo inovativnostjo presežejo morebitne slovnične omejitve in jezik prilagodijo svojim potrebam. Vsekakor bi bilo treba na tem področju opraviti še več podrobnejših analiz, da bi preverili, ali je kodno preklapljanje res povsem prosto in odvisno zgolj od domišljije govorcev ali pa vendarle obstajajo kakšne meje oziroma konstrukcije iz slovenskih in tujejezičnih prvin, ki niso mogoče.

V zvezi z integracijo preklapov v slovenščino smo opazovali dvoje: ortografsko in morfološko integracijo. Glede prve smo ugotovili, da je zapis velike večine (88 %) preklapov nepodomačen oz. prevzet citatno, pri čemer smo upoštevali samo jedro besede. Pri drugi smo opazovali, ali sta razvidna slovenska končnica in/ali obrazilo – takšnih preklapov je bilo dejansko še manj (8,3 %) kot tistih s podomačenim zapisom, med njimi pa je največ glagolov. Tudi na te rezultate so vplivali »neuvrščeni« preklopi, tvorjeni v skladu s slovnico tujega jezika, ter samostalniki z ničto končnico, glede katerih ostaja odprto vprašanje, ali bi jih uporabniki v drugačnem kontekstu pregibali »po slovensko«. Ko smo iz označenega korpusa izluščili preklope, ki imajo hkrati oznako za (delno) podomačenost zapisa in razvidno končnico in/ali obrazilo, smo ugotovili, da jih je skupno 50 oziroma 3,5 % vseh preklapov. Večina preklapov torej ni integriranih v slovenski jezik.

Med raziskavo smo posebno pozornost namenili kodnim preklpom v nemškem jeziku, saj se je že med označevanjem korpusa izkazalo, da je prav pri germanizmih problematično ločevanje med kodnimi preklopi in prevzetimi besedami zgolj na podlagi jezikovnih priročnikov. Izkazalo se je, da so nemški kodni preklopi v povprečju krajši, skoraj izključno znotrajstavčni, 80 % pa jih je enobesednih. V primerjavi z angleščino je bil njihov zapis v večji meri podomačen, prav tako je bilo med njimi več preklpov z razvidno končnico in/ali obrazilom. Vse to podpira našo dodatno hipotezo, da gre pri teh nemških preklpih večinoma za germanizme, ki so zaradi dolge zgodovine jezikovnih stikov med nemščino in slovenščino bolj integrirani v slovenščino ter so pravzaprav že na poti, da postanejo prevzete besede, vključene tudi v jezikovne priročnike, seveda z ustreznim kvalifikatorjem.

Za vsak tvit v korpusu Janes je bil že predhodno označen spol avtorja, zato smo se odločili preveriti, ali je mogoče zaznati kakšne razlike v preklapljanju med ženskami in moškimi. Izkazalo se je, da bistvenih razlik ni, s čimer smo potrdili izhodiščno hipotezo.

V kvalitativnem delu analize smo si ogledali, kakšne diskurzne funkcije lahko opravljajo kodni preklopi. Ugotovili smo, da tviteraši preklaplajo v najrazličnejših kontekstih, zato so tudi funkcije preklpov lahko zelo raznolike. Tujejezične prvine npr. omogočajo zgoščeno izražanje in gospodarno rabo jezika ter služijo kot dodatno pojasnilo, poudarek ali referenca na zunanje okoliščine. Z njimi tviteraši odgovarjajo sogovornikom, jih nagovarjajo z (retoričnimi) vprašanji, izražajo svoja čustva ali podkrepijo svoj odnos do napisanega.

Iz pregleda zaznanih tematskih polj lahko ugotovimo, da se preklopi velikokrat nanašajo na televizijske serije in filme. Tu velja izpostaviti našo izhodiščno odločitev, da naslovov skladb, filmov ipd. ne označujemo kot prekop, saj gre za imena, ki načeloma niti nimajo slovenskih ustreznice – če bi jih upoštevali, bi bilo še bolj očitno, da so preklopi pogosto povezani s kulturo, predvsem iz angleško govorečega prostora. Nadalje ne preseneča, da je kar nekaj tvitov povezanih z računalništvom oz. novimi tehnologijami, kjer slovenske ustreznice včasih potrebujejo več časa, da se uveljavijo, po drugi strani pa je med strokovnjaki sprejeta uporaba tujih terminov. Kodni preklopi se pojavljajo tudi v zvezi s prostočasnimi dejavnostmi, kot sta šport in (kulturno specifična) hrana, kar nekaj pa jih je povezanih s Twitterjem in zanj specifičnimi izrazi (npr. »handle« ali »RT«).

V zadnjem delu analize smo se dotaknili še frazeoloških enot, pri čemer smo samo izpostavili zanimive primere pregovorov, idiomov, fraznih glagolov ter stalnih besednih zvez in kolokacij. Na tem področju bi lahko našli dovolj gradiva za še eno magistrsko delo, vsekakor pa je že iz

našega kratkega pregleda razvidno, da tviteraši dobro poznajo angleščino in se znajo z njeno pomočjo izražati na zelo barvite načine. V zvezi s tem bi bilo zanimivo raziskati tudi interference, do katerih prihaja, ko slovenski govorci dobesedno prevedejo in uporabijo takšne frazeme v slovenščino.

Sklenemo lahko, da kodni preklopi ne služijo zgolj zapolnjevanju leksikalnih vrzeli, temveč dajejo avtorjem širše možnosti za izražanje želenega sporočila ter ustvarjanje svoje spletne identitete in sloga. V magistrskem delu se namenoma nismo ukvarjali z vrednostno presojo, ali tovrstna uporaba tujejezičnih prvin ogroža ali bogati slovenščino, saj bodo mnenja glede tega vedno deljena in odvisna od osebnih prepričanj. Kljub temu je takšna jezikovna raba realno dejstvo – in to ne zgolj na Twitterju ali spletnih družbenih omrežjih – zato si zasluži empirično obravnavo. Nenazadnje bi morali kodno preklapljanje najprej dobro razumeti, preden ga lahko obsojamo ali hvalimo. Poleg tega je mešanje jezikov pomembna tema tudi za področje računalniške obdelave naravnega jezika, strojnega prevajanja in drugih tehnologij, kjer morajo algoritmi znati prepoznati jezik, iz katerega prihaja posamezna beseda – to pa je toliko težje v nestandardnih, večjezičnih besedilih, ki so značilna za velik del računalniško posredovane komunikacije.

5 Summary

The internet has connected people all around the globe and significantly changed many aspects of our everyday lives. In linguistics, it presents a multilingual situation where languages come into contact on an unprecedented scale, which influences the ways how languages are used and shaped by the users. The English language plays a dominant role online, serving as a source of new vocabulary in other languages, which is why many see it as a threat, particularly for minor languages. One of the outcomes of such language contact is code-switching (hereafter CS) – the use of words from two or more languages within one sentence or utterance.

This thesis aims to explore CS in the tweets collected as part of the Janes corpus of Slovene computer-mediated communication. It is divided into two main parts – the first introduces the theoretical background and the second includes the empirical research and results.

The theoretical part first introduces the general characteristics of the language on the internet (Crystal 2006 and 2011, Baron 2008 etc.) and the language used on the social media and Twitter (Herring 2012, Page et al. 2014 etc.). The next chapter is dedicated to language contact (Paolillo 2007, Thomason 2001 etc.), focusing also on the influence of English on other languages, in particular Slovene (Šabec 2014a). Then, the related work on CS is presented (Gardner-Chloros 2009, Myers-Scotton 2002 etc.), including the definition and various types of CS (Blom and Gumperz 2000, Poplack 1980, Appel and Muysken 2005 etc.), as well as the studies on CS in computer-mediated communication (Sebba 2012, Androutsopoulos 2013 etc.). Then, the Matrix Language Framework model (Myers-Scotton 1997) is briefly introduced as well as some grammatical constraints governing CS proposed by other researchers (cf. Clyne 2009), and various discourse functions of CS (Androutsopoulos 2013, Mikolič Južnič and Čibej 2016).

The last chapter in the theoretical part introduces the Janes corpus, developed as part of the national research project JANES – *Resources, Tools and Methods for the Research of Non-standard Internet Slovene*. The version 0.4 of the corpus includes 175 million words and is divided into five sub-corpora (containing tweets, forum posts, blog texts, and comments on news articles and on Wikipedia pages and users). The corpus of tweets is the largest, including 90 million words, and the meta data of each tweet include the author's gender and the level of standardness (Fišer et al. 2016).

Prior to conducting our research, we came up with the following hypotheses:

- 1) the non-standard texts will include more CS than the standard texts;
- 2) the majority of code-switches will be in English but other languages will also be present

(e.g. German, Spanish, Serbo-Croatian);

- 3) there will be a higher proportion of intrasentential code-switches than intersentential code-switches, and intrasentential code-switches will be, on average, shorter than intersentential code-switches;
- 4) there will be a higher proportion of open-class words (nouns, verbs, adjectives) than closed-class words (prepositions, conjunctions, pronouns);
- 5) the majority of code-switches will not be orthographically integrated (e.g. they will be written in accordance with the norm of the source language);
- 6) the majority of code-switches will not be morphologically integrated, but verbs will be more often integrated (e.g. with Slovenian endings) than nouns and adjectives;
- 7) there will be no significant differences in CS in relation to the author's gender;
- 8) code-switches will be used to fill lexical gaps or to achieve special emphasis;
- 9) in terms of the topic, CS will be often related to IT, popular culture and culturally specific terms;
- 10) multi-word code-switches will also include idioms and collocations.

The empirical part was further divided into two sections, namely the quantitative and qualitative analysis. For the research, a corpus of Slovenian tweets had to be prepared and annotated, which was done in two steps. First, a general sample was created, which included 1600 tweets with at least one foreign word, of which there was an equal number of standard and non-standard texts. However, during the annotation process many false positives were discovered, e.g. tweets without any CS (either because they were written entirely in a foreign language or due to the errors resulting from the automatic recognition of foreign words). Therefore, another sample of 1600 tweets was prepared, which only included non-standard texts.

Prior to the annotation, which was carried out in WebAnno, the guidelines for annotating CS (Appendix A) were determined. It was particularly important to decide on how to separate code-switched items from borrowings/loanwords. We chose the same approach as Deuchar 2006 and Rosenberg 2013, i.e. if a word was listed in one of the established Slovenian dictionaries, it was considered a borrowing, not a code-switch. In both samples, non-relevant tweets were deleted and each code-switch was manually annotated on five levels: language (English, German etc.), type of CS (intra- and intersentential), orthography (assimilated, partially assimilated or non-assimilated), morphology (not evident or evident from the ending/affix) and part of speech (noun phrase, verb phrase etc.). The annotated tweets were exported in the .xlsx format so the analysis could be performed in Excel.

In addition to the quantitative analysis (number of code-switches, proportion of languages etc.), a qualitative study was also carried out by looking at semantic fields in which CS occurs as well as the discourse functions of CS. Any interesting observations were noted in the Excel file next to individual tweets. The purpose was to highlight the variety of situations in which CS occurs in Slovenian tweets, while the results may serve as a basis for a more detailed study in the future.

The overall results showed that CS is not a rare phenomenon in the tweets written by Slovenian users, although its frequency in Slovenian tweets in general cannot be estimated. Our sample also showed great variety in terms of different contexts and functions performed by CS. In the corpus, 67% of code-switches were intrasentential and 33% intersentential, consisting on average of 1.26 words. About one half of all code-switches were single-word switches, which confirmed the hypothesis and findings from the theory, supporting the claim that CS can also be used by people who do not have a vast knowledge of the foreign language. Many of the single-word switches were abbreviations such as *lol* or *omg*, typical for online communication. It was also proven that CS occurs more often in non-standard texts.

Among the languages found in the corpus, English clearly dominates (90%), followed by Serbo-Croatian (4.6%) and German (3.3%). In total, 9 different foreign languages were detected, however, the number of examples from these languages was very low and they were mostly formulaic expressions and culturally specific terms. Such a large share of the English language can be attributed to many reasons: especially younger generations learn(ed) English as a second language in school, it is widely present in popular culture and it extends its influence as the main language on the internet.

The hypothesis regarding the word class was confirmed only partially – the proportion of code-switched noun phrases was quite high (33%), but the share of adjectival and verb phrases was lower (9% and 5%, respectively). The share of interjections (15%) was somewhat surprising, but can be explained to some extent by the frequent use of the above-mentioned abbreviations. Due to the lack of data, only one grammatical constraint in Slovenian-English CS could be (partially) tested: the constraint on the switchability of closed-class words does not apply since particles, conjunctions and pronouns had been switched. However, there were only a few such cases and a more detailed study would be needed to properly determine any valid constraints on CS in the Slovenian language.

Furthermore, the data showed that the majority of code-switches was not orthographically or

morphologically integrated; instead, they were formed in accordance with the grammar of the embedded language (in most cases English). An interesting exception were German code-switches, which were mostly single-word items and assimilated to a greater degree. This is probably due to the fact that the Slovenians had a long history of contacts with the German-speaking areas, including the neighbouring Austria, so there are already many German borrowings listed in the Slovenian dictionaries. Some of the discovered code-switches are therefore probably on their way to becoming established borrowings.

The hypothesis about the gender was confirmed as well since no significant differences in CS between male and female authors were found.

The qualitative analysis showed that CS fulfils many different functions: foreign words can be used to achieve language economy, add emphasis or a reference, address another person or express emotions. In terms of the semantic fields, we found that CS was related to popular culture, in particular TV shows, sport, (culturally specific) food and Twitter. All of this was not surprising as the English language is dominant in these areas, so people have easier access to the related vocabulary. It was also noted that many code-switches were in fact idiomatic expressions, phrasal verbs, collocations and even proverbs, which shows that the authors are quite competent in English and can use foreign languages to better express themselves in a given situation.

To conclude, the research showed that CS is not used only to fill in the lexical gaps. Indeed, Twitter users use foreign words and phrases for a variety of purposes, which allows them to convey their message in the best possible way. In the thesis, the question of whether such use of foreign languages, especially English, threatens or enriches the Slovenian language was deliberately avoided. As is evident from the findings, CS is definitely part of our communication – not just on Twitter or in other social media – so it should be empirically examined before any judgements are made. Additionally, code mixing is an important issue in natural language processing, machine translation and other language technologies where algorithms should first decipher the language of each word – a task made even more difficult in non-standard, multilingual texts, which are especially common in computer-mediated communication.

6 Viri in literatura

- Androutsopoulos, Jannis, 2006: Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics* 10/4. 419–438.
- Androutsopoulos, Jannis, 2013a: Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism* 0/0. 1–21.
- Androutsopoulos, Jannis, 2013b: Code-switching in computer-mediated communication. Herring, Susan C., Dieter Stein, Tuija Virtanen (ur.): *Handbook of the Pragmatics of CMC*. Berlin: Mouton de Gruyter. 667–94.
- Appel, Rene in Pieter Muysken, 2005: *Language Contact and Bilingualism*. Amsterdam: Amsterdam University Press.
- Auer, Peter, 1984: Bilingual conversation. Amsterdam, Netherlands: John Benjamins.
- Baron, Naomi S., 1984: Computer mediated communication as a force in language change. *Visible Language*, XVIII 2, 118–41.
- Baron, Naomi S., 2003: Language of the Internet. Farghali, Ali (ur.): *The Stanford Handbook for Language Engineers* (pp.). Stanford: CSLI Publications. 59–127.
- Baron, Naomi S., 2008: *Always on: Language in an Online and Mobile World*. Oxford; New York: Oxford University Press.
- Baron, Naomi, 2000: *From Alphabet to E-mail: How Written English Evolved and Where It's Headed*. London: Routledge.
- Blom, Jan-Petter in John J. Gumperz, 2000: Social meaning in linguistic structure: Codeswitching in Norway. Wei, Li (ur.): *Bilingualism reader*. London; New York: Routledge. 102–126.
- Boztepe, Erman, 2003: Issues in Code-Switching: Competing Theories and Models. *Working Papers in TESOL & Applied Linguistics*, 3 (2), 1–27.
- Clyne, Michael, 2009: Constraints on code switching: how universal are they? *Linguistics*, 25 (4). 739–764.
- Crystal, David, 2006: *Language and the internet*. New York: Cambridge University Press.
- Crystal, David, 2011: *Internet linguistics: a student guide*. Abingdon; New York: Routledge.
- Danet, Brenda in Susan C. Herring, 2007: Introduction: Welcome to the Multilingual Internet. Danet, Brenda in Susan C. Herring (ur.): *The multilingual internet: language, culture, and communication online*. Oxford; New York: Oxford University Press. 3–39.
- Deuchar, Margaret, 2006: Welsh-English code-switching and the Matrix Language Frame model. *Lingua* 116. 1986–2011.
- DiNucci, Darcy, 1999: Fragmented future. Print Magazine, 4, 32. http://darcyd.com/fragmented_future.pdf/. (Dostop 10. 8. 2017)

- Eastman, Carol M., 1992: Codeswitching as an urban language contact phenomenon. *Journal of Multilingual and Multicultural Development*, 13. 1–17.
- Eisenstein, Jacob, 2013: What to do about bad language on the internet. *Proceedings of NAACL-HLT 2013*, Atlanta, Georgia, 9.–14. junij 2013. 359–369.
- Erjavec, Tomaž, Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Darja Fišer, Cyprian Laskowski, Katja Zupan, 2016: Annotating CLARIN.SI TEI corpora with WebAnno. *Proceedings of the CLARIN Annual Conference 2016*. Aix-en-Provence, Francija.
- Fišer, Darja, 2016: Slovenščina 2.0: »Računalniško posredovana komunikacija«. *Slovenščina 2.0*, 4 (2): i–iv, <http://slovenscina2.0.trojina.si/arhiv/2016-2/>. (Dostop 10. 8. 2017)
- Fišer, Darja, Tomaž Erjavec, Nikola Ljubešić, 2016. JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0*, 4 (2): 67–99.
- Gardner-Chloros, Penelope in Daniel Weston, 2015: Code-switching and multilingualism in literature. *Language and Literature* 24/3. 182–193.
- Gardner-Chloros, Penelope, 2009: *Code-switching*. Cambridge (UK); New York: Cambridge University Press.
- Gardner-Chloros, Penelope, 2010: Contact and Code-Switching. Hickey, Raymond (ur.): *The Handbook of Language Contact*. Blackwell Publishing. 188–207.
- Georgakopoulou, Alexandra in Tereza Spilioti, 2016: Introduction. Georgakopoulou, Alexandra in Spilioti, Tereza (ur.): *The Routledge handbook of language and digital communication*. Abingdon; New York: Routledge. 1–15.
- Gerrand, Peter, 2007: Estimating Linguistic Diversity on the Internet: A Taxonomy to Avoid Pitfalls and Paradoxes. *Journal of Computer-Mediated Communication* 12/4. 1298–1321.
- Grosjean, François, 1982: *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Pres.
- Gumperz, John J., 1982: *Discourse strategies*. Cambridge: Cambridge University Press.
- Haugen, Einar, 1950: The analysis of linguistic borrowing. *Language* 26 (2). 211–231.
- Herring, Susan C., 2001: Computer-mediated Discourse. Schiffrin, Deborah, Deborah Tannen in Heidi E. Hamilton (ur.): *The Handbook of Discourse Analysis*. Malden; Oxford: Blackwell Publishers. 613–634.
- Herring, Susan, 1996: *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam; Philadelphia: J. Benjamins Publishing Company.
- Herring, Susan, 2012 (in press): Discourse in Web 2.0: Familiar, Reconfigured, and Emergent. Tannen, Deborah in Anna Marie Trester (ur.): *Discourse 2.0: Language and New Media*. Washington, DC: Georgetown University Press. <http://info.ils.indiana.edu/~herring/-GURT.2011.prepub.pdf/>. (Dostop 22. 8. 2017)
- Hughes, Claire E., Elizabeth S. Shaunessy, Alejandro R. Brice, Mary Anne Ratliff, Patricia Alvarez MyHatton, 2006: Code Switching among Bilingual and Limited English Proficient

Students: Possible Indicators of Giftedness. *Journal for the Education of the Gifted*, 30/1. 7–28.

Internet World Stats. <http://www.internetworldstats.com/stats7.htm/>. (Dostop 20. 8. 2017)

Jakop, Nataša, 2006: *Pragmatična frazeologija*. Ljubljana: Založba ZRC

Joshi, Aravind K., 1982: Processing Of Sentences With Intra-Sentential Code-Switching. *Proceedings of the 9th conference on Computational linguistics* – Volume 1. 145–150.

Kachru, Braj B., 1983: On mixing. Kachru, Braj (ur.): *The Indianization of English: The English language in India*. New Delhi: Oxford University Press. 193–207

Kranjc, Simona, 2003: *Jezik v elektronskih medijih*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete.

Krošelj, Andreja, 2015: Code-switching in comments on Slovene Internet journalistic articles. Graduation thesis. Univerza v Mariboru. <https://dk.um.si/IzpisGradiva.php?id=48261/>. (Dostop 23. 8. 2017)

Laslo, Ksenija, 2016: *Vpliv angleščine na slovenščino na družbenih omrežjih: slovnica in pravopis*. Diplomsko delo. Univerza v Mariboru. <https://dk.um.si/IzpisGradiva.php?id=59179/>. (Dostop 20. 8. 2017)

Lenihan, Aoife, 2011: “Join Our Community of Translators”: Language Ideologies and/in Facebook. Thurlow, Crispin in Kristine Mroczek (ur.): *Digital discourse: language in the new media*. Oxford [etc.].

Ljubešić, Nikola, Darja Fišer, Tomaž Erjavec, Jaka Čibej, Dafne Marko, Senja Pollak, Iza Škrjanec, 2015. Predicting the level of text standardness in user-generated content. *Proceedings of RANLP'15 Conference*, 7.–9. september 2015, Hissar, Bolgarija. 371–378.

Lončarić, Jelena, 2014: *Croatian-English Code-Switching Patterns of Croatian Facebook Users*. Master's Thesis. University of Zagreb. <http://darhiv.ffzg.unizg.hr/5090/1/J-elena%20Loncaric%20-%20diplomski.pdf/>. (Dostop 23. 8. 2017)

Markelj, Klemen, 2015: *Anglicizmi v spletnem oglaševanju na Facebooku*. Magistrsko delo. Univerza v Mariboru. <https://dk.um.si/IzpisGradiva.php?id=47376/>. (Dostop 20. 8. 2017)

Marko, Dafne, 2016: The Use of Alphanumeric Symbols in Slovene Tweets. *Proceedings of the 4th Conference on CMC and Social Media Corpora for the Humanities*, Ljubljana, Slovenia, 27.–28. september 2016. 48–53.

Michelizza, Mija, 2007: Splèt stereotipov o spletni slovenščini. Novak-Popov, Irena (ur.): *Stereotipi v slovenskem jeziku, literaturi in kulturi: zbornik predavanj/43. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik. 206–210.

Mikolič Južnič, Tamara in Jaka Čibej, 2016: Slovenščina v stiku z drugimi jeziki. *Poletni raziskovalni tabor JANES 2016*. <http://nl.ijs.si/janes/wp-content/uploads/2016/01/tabor-cibej-predavanje.pdf/>. (Dostop 20. 8. 2017)

- Muysken, Pieter, 2010: Scenarios for Language Contact. Hickey, Raymond (ur.): *The Handbook of Language Contact*. Blackwell Publishing. 263–281.
- Myers-Scotton, Carol in Janice L. Jake, 2000: Matching lemmas in a bilingual language competence and production model: evidence from intrasentential code-switching. Wei, Li (ur.): *Bilingualism reader*. London; New York: Routledge. 265–302.
- Myers-Scotton, Carol in William Ury, 1977: Bilingual Strategies: The Social Functions of Code-Switching. *International Journal of the Sociology of Language* 13. 5–20
- Myers-Scotton, Carol, 1997: *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.
- Myers-Scotton, Carol, 2000: Code-switching as indexical of social negotiations. Wei, Li (ur.): *Bilingualism reader*. London; New York: Routledge. 127–153.
- Myers-Scotton, Carol, 2002: *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford; New York: Oxford University Press.
- Myllärinen, Essi, 2014: "Mä koitan olla kutsumatta niit infectedeiks koska anglismi" - Code-switching in Finnish online gaming videos. Master's Thesis. University of Jyväskylä. <https://jyx.jyu.fi/dspace/handle/123456789/44787/>. (Dostop 23. 8. 2017)
- Page, Ruth, 2014: Hoaxes, hacking and humour: analysing impersonated identity on social network sites. Seargeant, Philip in Caroline Tagg (ur.): *The Language of Social Media: Identity and Community on the Internet*. Basingstoke: Palgrave Macmillan. 46–64.
- Page, Ruth, David Barton, Johann W. Unger, Michele Zappavigna, 2014: *Researching language and social media: a student guide*. London; New York: Routledge.
- Paolillo, John C., 1996: Language choice on soc.culture.punjab. *Electronic Journal of Communication*, 6(3). <http://www.cios.org/EJCPUBLIC/006/3/006312.HTML/>. (Dostop 20. 8. 2017)
- Paolillo, John C., 1999: The Virtual Speech Community: Social Network and Language Variation on IRC. *Journal of Computer-Mediated Communication* 4/4.
- Paolillo, John C., 2007: How Much Multilingualism? Language Diversity on the Internet. Danet, Brenda in Susan C. Herring (ur.): *The multilingual internet: language, culture, and communication online*. Oxford; New York: Oxford University Press. 408–430.
- Poplack, Shana in Nathalie Dion, 2012: Myths and facts about loanword development. *Language Variation and Change*, 24. 279–315.
- Poplack, Shana, 1980: Sometimes I'll start a sentence in Spanish y termina en español: toward a typology of code-switching. *Linguistics* 18(7/8): 581–618.
- Poplack, Shana, 2015: Code Switching: Linguistic. *International Encyclopedia of the Social and Behavioral Sciences*, 2nd edition. Elsevier Science Ltd. 918–925.
- Poplack, Shana, David Sankoff in Christopher Miller, 1988: The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26: 47–104.

- Rosenberg, Johanna, 2013: *"I chuckled. Rolig pun där."* A grammatical investigation of Swedish-English code-switching in two web discussion forums. Lund University.
- Sankoff, David in Shana Poplack, 1981: A formal grammar for code switching. *Papers in Linguistics*. 14 (1). 3–46.
- Sankoff, Gillian, 2001: Linguistic Outcomes of Language Contact. <http://www.ling.u-penn.edu/~gillian/Interlang.doc.pdf/>. (Dostop 22. 8. 2017)
- Savić, Jelena M., 1996: *Code-switching: theoretical and methodological issues*. Beograd: College of philology.
- Seargeant, Philip in Caroline Tagg, 2014: Introduction: The language of social media. Seargeant, Philip in Caroline Tagg (ur.): *The Language of Social Media: Identity and Community on the Internet*. Basingstoke: Palgrave Macmillan. 1–20.
- Sebba, Mark, 2012: Multilingualism in written discourse: An approach to the analysis of multilingual texts. *International Journal of Bilingualism* 17/1. 97–118.
- Sherman, Todd, 2017: Coming soon: express even more in 140 characters. *Blog družbe Twitter*. https://blog.twitter.com/official/en_us/a/express-even-more-in-140-characters.html/. (Dostop 29. 8. 2017)
- Singh, Rajendra, 1985: Grammatical constraints on code-switching: Evidence from Hindi-English. *Canadian Journal of Linguistics*, 30. 33–45.
- Smernice za označevanje korpusa slovenskih tvitov JANES: tokenizacija, stavčna segmentacija, normalizacija, lematizacija in oblikoskladenjsko označevanje*, 2016. <http://nl.ijs.si/janes/wp-content/uploads/2014/09/JANES-jezikoslovne-smernice-v0.9.pdf/>. (Dostop 22. 8. 2017)
- Snoj, Marko, 2005: O prevzetih besedah, tujkah in izposojenkah. *Slovenski jezik – Slovene linguistic studies* 5. 113–122.
- Sridhar, Shikaripur N. in Sridhar, Kamal, 1980: The syntax and psycholinguistics of bilingual code-mixing. *Canadian Journal of Psychology*, 34. 407–416.
- Stabej, Marko, Helena Dobrovoljc, Simon Krek, Polona Gantar, Damjan Popič, Špela Arhar Holdt, Darja Fišer, Marko Robnik Šikonja, M. (2016): Slovenščina Janes: pogovorna, nestandardna, spletna ali spretna? *Slovenščina 2.0*, 4 (2): 100–126.
- Šabec, Nada, 2009: Sloglish or the mixing/switching of Slovene and English in Slovene blogs. *Slavia Centralis* 2/1. 32–42.
- Šabec, Nada, 2014a: Raba slovenščine in angleščine v fizičnem in virtualnem prostoru. *Prostor v jeziku in jezik v prostoru* 62/3. 395–409.
- Šabec, Nada, 2014b: Blagovna znamka/brand, branding/znamčenje --- : vpliv angleščine na sodobno slovensko poslovno terminologijo. Warchał, Mateusz (ur.): *W podróży za słowem: księga pamiątkowa z okazji Jubileuszu 70-lecia urodzin profesora Emila Tokarza*. Bielsko-Biała: ATH. 167–178.

Šekli, Matej, 2015: Pomenska polja nemških izposojenk v slovenščini. *Jezikoslovni zapiski*, letnik 21, številka 2, str. 31–44.

Štefančič, Natalija, 2016: *The influence of English on Slovene in advertisements in the field of health*. Graduation thesis. Univerza v Mariboru. <https://dk.um.si/IzpisGradiva.php?id=62957/>. (Dostop 20. 8. 2017)

Tagg, Caroline in Philip Seargeant, 2014: Audience design and language choice in the construction and maintenance of translocal communities on social network sites. Seargeant, Philip in Caroline Tagg (ur.): *The Language of Social Media: Identity and Community on the Internet*. Basingstoke: Palgrave Macmillan. 161–185.

Tastan, Feride Duygu, 2012: *Language Contact in Computer-Mediated Communication: The Use of English on Twitter by Turkish Native Speakers*. Trabajo fin de Máster. Universidad Complutense Madrid. <http://eprints.ucm.es/15382/>. (Dostop 23. 8. 2017)

Thomason, Sarah G., 2001: *Language contact: an introduction*. Edinburgh: Edinburgh University Press.

Thurlow, Crispin (in press/2017): Digital discourse: Locating language in new/social media. Burgess, Jean, Alice E. Marwick in Thomas Poell (ur.), *The SAGE Handbook of Social Media*. New York: Sage. http://www.academia.edu/31483202/Digital_discourse_Locating_language_in_new_social_media/. (Dostop 2. 8. 2017)

Thurlow, Crispin in Kristine Mroczek, 2011: Introduction: Fresh perspectives on new media sociolinguistics. Thurlow, Crispin in Kristine Mroczek (ur.): *Digital discourse: language in the new media*. Oxford [etc.]. xix-xliv.

Toporišič, Jože, 1992: *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.

Toporišič, Jože, 2004: *Slovenska slovnica*. Maribor: Obzorja.

W3Techs. https://w3techs.com/technologies/overview/content_language/all/. (Dostop 20. 8. 2017)

Wei, Li, 2000a: Dimensions of bilingualism. Wei, Li (ur.): *Bilingualism reader*. London; New York: Routledge. 2–21.

Wei, Li, 2000b: Introduction to Part One. Wei, Li (ur.): *Bilingualism reader*. London; New York: Routledge. 54–56.

Weinreich, Uriel, 1953: *Languages in contact, findings and problems*. New York: Linguistic Circle of New York.

Zappavigna, Michele, 2012: *Discourse of Twitter and social media: [how we use language to create affiliation on the web]*. London [etc.]: Bloomsbury.

Zappavigna, Michele, 2014: CoffeeTweets: bonding around the bean on Twitter. Seargeant, Philip in Caroline Tagg (ur.): *The Language of Social Media: Identity and Community on the Internet*. Basingstoke: Palgrave Macmillan. 139–160.

Zwitter Vitez, Ana in Darja Fišer, 2015. Elementi interakcije v govorjenih in spletnih besedilih. Fišer, Darja (ur.). *Zbornik konference Slovenščina na spletu in v novih medijih*. Ljubljana: Znanstvena založba Filozofske fakultete, 87–90.

7 Priloge

7.1 Priloga A: Smernice za označevanje kodnega preklapljanja v korpusu slovenskih tvitov Janes

Kazalo vsebine

- 1 Splošna načela
- 2 Stavčna segmentacija
- 3 Večbesedni preklopi
- 4 Jezik
- 5 Zapis
- 6 Morfologija
- 7 Besedna vrsta
- 8 Oznake za označevanje in njihov pomen

Smernice podajajo jezikoslovna načela za označevanje lastnosti kodnih preklapov oziroma tujejezičnih besed v slovenskih tvitih.

1 Splošna načela

1. Označujemo samo tujejezične besede v tvitih, v katerih je uporabljena tudi slovenščina. Pri presoji, ali je določena beseda slovenska, uporabimo načelo vključenosti v Slovar slovenskega knjižnega jezika, Slovenski pravopis ali Sloleks – če besedo najdemo v katerem od teh priročnikov, jo štejemo za slovensko. Podrobneje v 4. razdelku Jezik.
2. Če je tweet v celoti v tujem jeziku oz. več tujih jezikih, avtomatsko generiran ali popolnoma nerazumljiv, tweet izberemo. V takšnem tweetu ne označujemo ničesar drugega.
3. Pri označevanju stavčnih mej upoštevamo *Smernice za označevanje korpusa slovenskih tvitov JANES* (v nadaljevanju Osnovne smernice). Relevantna pravila so vključena v naslednjem razdelku.
4. Na prvem nivoju vsakemu preklopu pripišemo jezik in vrsto preklopa. Na drugem nivoju za vsako besedno zvezo (gl. 3. razdelek Večbesedni preklopi) označimo zapis, morfologijo in besedno vrsto. Večinoma se oba nivoja prekrivata, torej se oznake v WebAnnu raztezajo čez isti sklop besed; do razlike prihaja pri določenih večbesednih preklapih, kar je podrobneje pojasnjeno v nadaljevanju.
5. Kadar se tujejezična prvina pojavi kot hešteg oziroma kot del heštega, kodni preklop dodatno označimo z oznako hashtag.
6. Seznam vseh oznak za označevanje je vključen v 8. razdelku.

Zapis ponazoritvenih primerov:

Primeri so zapisani ležeče in v narekovajih, npr. »*fair enough*«. Pri večbesednih preklapih oz. več ločenih zaporednih preklapih so skupaj z neprekinjeno črto podčrtane besede, ki jih v WebAnnu označimo z eno skupno oznako, če pa je med njimi nepodčrtan presledek, to pomeni, da gre za dve ločeni enoti.

2 Stavčna segmentacija¹

Cilj: Besedila so pravilno ločena na stavke,² konec vsakega stavka pa je označen.

Smernice:

1. V celotnem tvitu preverimo, ali je avtomatska stavčna segmentacija pravilna.
2. Če del tvita deluje kot samostojen stavek, ga tako tudi obravnavamo³ (»@tviteraš Najbrž ne . ¶ :) ¶ Kot rečeno : bolje BO . ¶ Zrihtamo , ko utegnemo . ¶ (PS : tudi v veselje " konkurence " ;)«).
3. Merilo za konec stavka je predvsem ločilo, ki deluje kot končno v stavku, npr. pika, klicaj, vprašaj, narekovaj ali večpičje (»Kaj praviš ? ¶ Aha !«).
4. Če ni dobrega razloga, da nekaj obravnavamo kot dva stavka, naj ostane eden (»@tviteraš pri meni naloži CSS .. kar pa ne pomeni , da stran zgleda lepo :)« → en stavek, ker večpičje deluje bolj kot vejica, ne kot pika).
5. Konec tvita je avtomatično tudi konec stavka, zato tega ne označujemo.

Težavni primeri:

1. Večpičje
 - a. Ponavadi je končno ločilo (»@tviteraš po moje se jo izogiba kot hudič križa. ¶ Glavn da on spet laja ... ¶ :-)))))«).
 - b. Včasih označuje zgolj elipso ali zamolk sredi stavka – v takšnem primeru ni končno ločilo (»To se mi zdi ... neumno.«).
2. Imena (@ime)
 - a. Če se pojavljajo sredi stavka, so del stavka (»neka baka :) uleti pa praša če loh gre kr naprej«, »sej #tarca je pa dons kr ok«, »sej je rekla @tviteraš d je treba to drgac«).
 - b. Če se pojavljajo na začetku ali koncu tvita, jih ne označujemo oz. upoštevamo pri določanju, ali je preklap znotraj- ali medstavčni (npr. v primeru »@tviteraš Lol. Raj za pivo zber :P« besedo »lol« označimo kot medstavčni preklap, ime pa ignoriramo).
3. Emotikoni
 - a. Če nadomeščajo končno ločilo, zaznamujejo konec stavka (»kot da je to važn :)) ¶ nobenga to ne briga vec sploh«).
 - b. Če sledijo končnemu ločilu, jih obravnavamo kot samostojen stavek (»Sonce, sneg in pot pod noge! ¶ :) ¶ Gremo v hribe!«).
4. Znak *
 - a. Kot mejo stavka označimo tudi znak * v primerih, kot je »*starts thinking * *hurts just a little bit *« (torej gre za medstavčni preklap).

¹ Povzeto in prilagojeno po: <http://nl.ijs.si/janes/wp-content/uploads/2014/09/JANES-jezikoslovne-smernice-v0.9.pdf>

² "Stavek" uporabljamo v širšem pomenu, torej kot najmanjšo samostojno enoto jezikovnega sporočila, ki bi se v standardni slovenščini začela z veliko začetnico in zaključila s končnim ločilom ("poved" v slovenistični terminologiji).

³ V smernicah konec stavka za lažjo predstavo označujemo s simbolom ¶.

3 Večbesedni preklopi

Cilj: Pri vsakem kodnem preklopu sta označena jezik in vrsta preklopa (znotraj- ali medstavčni), pri posameznih besednih zvezah pa označimo še obliko zapisa, morfologijo in besedno vrsto.

Smernice:

1. Pri večbesednih preklapih skupno (tj. z eno oznako) označimo JEZIK in VRSTO PREKLOPA. Za seznam znak in njihov pomen gl. tabelo na koncu Smernic.
2. Za vsako besedno zvezo posebej označimo ZAPIS, MORFOLOGIJO in BESEDNO VRSTO, pri čemer za posamezno besedno zvezo štejemo:
 - a. **stavek ali daljši sklop besed, ki jim ni mogoče določiti ene besedne vrste (N)**
 - i. »@tviteraš I need this so bad right now . Ne mačke . Svaljkanje« (MS)
 - ii. »@tviteraš ohhh i know predobr kako je po 72 h urah nespanja ...« (ZS)
 - b. **ključnik (hashtag)**
 - i. »Tisto , ko si pred tvojiim komadom fentas koleno :S #fml #jobparty« (označimo skupaj MS+ANG, nato zapis + morfologijo + bes. vrsto za vsak hešteg posebej)
 - c. **sklop ponovljenih besed:** »omg omg omg« ali »and and and« ali »jp, jp«
 - d. **samostalniško zvezo** skupaj s prilastkom (premodifier ali postmodifier)
 - i. pridevnik + samostalnik: »old boysov«, »scary shit«
 - ii. samostalnik + samostalnik: »buffalo burger«
 - iii. daljši prilastek: »take your kid to the work day«; »' i dont give a fuck ' attitude«
 - iv. samostalnik + PostM: »a breath of fresh air«
 - v. predlog + samostalnik = predložna zveza
 - e. **pridevniško zvezo** skupaj s prilastkom
 - i. pridevnik + prislov: »fair enough«
 - ii. prislov + pridevnik: »#oddlysatisfying«
 - f. **predložno zvezo:** »on a regular basis«
 - g. pri drugih kombinacijah zapis, morfologijo in besedno vrsto označimo za vsako besedo oziroma besedno vrsto posebej
 - i. »sej pravm , francozi ownajo horrorland .« (glagol + samostalnik)
 - ii. »awwwww kjuuuu si napisala ^^^« (medmet + pridevnik)

Posebnosti pri določanju meje oz. vrste preklopa:

1. Upoštevanje meje stavka
 - a. Če je med znotrajstavčnim preklpom in medstavčnim preklpom oziroma dvema znotrajstavčnima preklpomoma meja stavka, takšne preklope označimo vsakega zase.
 - i. »Moj HTC Cha Cha umira in spet bom pristala na tipkanju po touch screenu . ¶ I hate that .«
 - ii. »Jočem . @tviteraš there will be blood ! ¶ ;) btw jaz se kar pustim čakati za tisto kavo , čaj , pivo ... ;)«
 - iii. »adejno , a zdej še tviterji težijo z add your birthday .. ¶ kinky , lubčki , kinky ...«
 - b. Stavčne meje pa ne upoštevamo med dvema ali več medstavčnimi preklpi, ki jih na tem nivoju označimo skupno kot en medstavčni preklop.
 - i. »Pa smo tam. http://t.co/BvNxgwKiDH. 10min zamude. Evo je. Lušna punca. Tri punce. Fast driver. ¶ Smells like teen spirit. #romanjevljubljano«
2. Preklope znotraj heštegov, kjer je več besed napisanih skupaj, označimo, kot da bi bili zapisani razvezano (saj v WebbAnnu mogoče označiti samo dela besede).
 - ii. Primere, kot je »#dojenjemyass«, označimo kot znotrajstavčni preklop.
 - iii. Primere, kot je »#ilovemyjobs«, označimo kot medstavčni preklop.

4 Jezik

Cilj: Vsak kodni preklomp ima pripisano oznako za jezik, iz katerega je.

Smernice:

1. Pri določanju jezika smiselno upoštevamo Osnovne smernice v delu, ki se nanaša na tujejezične prvine. Če v tujejezičnem kontekstu nastopa beseda, ki bi jo lahko interpretirali kot slovensko, štejemo, da je del prevladujočega jezika v tvitu, torej ni preklomp
- a. V primeru »*Dobro jutro uz pjesmu " Dobro jutro*« vse skupaj upoštevamo kot HBS, torej tvit ni relevanten za raziskavo in ga izbrišemo.
- b. Primer »*@tviteraš Read my lips : p-f-f-f-f*.« upoštevamo kot v celoti angleški.
2. Smiselno enako velja za besedne zveze, kjer je en element sicer v SSKJ/SP/Sloleks, vendar je glede na kontekst treba besedno zvezo razumeti kot zaključeno enoto: *buffalo burger, blog post, party pooper, scam email, usb hubi* (vse angleške besedne zveze)
3. Latinsko okrajšavo VS (vs., v.) označimo kot ANG.
4. Medmeti: tudi tu upoštevamo načelo vključenosti v SSKJ/SP/Sloleks.
- a. Primeri slovenskih: *uf, eh, hm, pst, ah, mhm, no, vau, vav, mnja, jao, grr*
- b. Primeri angleških: *nah, ugh, ouje, meh, pfft, khm, ahm, uhm, blah, erm, ehm, aw, huh*
5. Tujejezične prvine in primeri, ki jih **ne** označujemo kot kodni preklomp:
 - a. **Besede, ki niso v SSKJ/SP/Sloleks, a jim ni mogoče določiti izvora:** *jupi, orng (orenk)*
 - b. **Neologizmi:** *biznisira, slochi*
 - c. **Imena,** npr. *medijev, mest, glasbenih skupin, lokalov, prirediteljev, športnih tekmovanj*
 - i. Izjema so države, saj imajo slovensko ustreznico:
 - ii. Če se ime pojavi znotraj preklopa, ga pustimo označenega kot del preklopa. V primeru »*too much Bieber all around*« denimo celotno zvezo označimo kot en znotrajstavčni preklomp (ne kot »*too much Bieber all around*«).
 - d. **Avtomatski tviti oz. deli tvitov**
 - i. Samodejno generirani predlogi pri deljenju strani: *via @youtube/@RevijaReporter*
 - ii. Besedila, generirana v različnih aplikacijah ali na spletnih straneh, npr. »*I'm at @7_Burger_kamnik in Kamnik https://t.co/NdJbAPdAqA*« = Swarmapp; »*Just completed a 10.71 km run*« = Runkeeper; »*Drinking an Asahi Super Dry @ Shambala*« = Untappd.com
 - e. **Naslovi pesmi, člankov, spletnih strani**
 - i. »*In zdonela je iz zvočnikov :)) Iron Maiden - Phantom of the Opera http...*«
 - ii. »*Dobro jutro vrtničarji ! Se zbira ideje za novo sezono ? Še ena : An Urban Micro-Farm in a Mobile Box http://t.co/UY0D11Ieo2 via @zite*«
 - f. **V celoti tujejezični tvit oz. preklopi med tujimi jeziki** (tviti brez slovenščine niso relevantni)
 - i. Tvite kot je »*Breaking : Diego Costa zove Marijanu . Vрати pare !*« izbrišemo.
 - g. **Tujejezične prvine v citiranem tvitu**
 - i. »*No , tok da veste . Ne se zajebavat z mano . " @conspiracyimage : Fact http://t... "*«
 - ii. »*Lej, @tviteraš, to je pa čist zate* . RT @FascinatingVids: How to make your own turtle« Tukaj besedo RT označimo kot preklomp, saj jo uporabnik napiše sam, drugi del pa ni relevanten, saj je citiran tuj avtor, zato ga pustimo neoznačenega.
 - h. **Interference na nivoju skladnje:** »*Rešitelj od Svete trojice ?*« ali »*tut t-2 je bl shit v mojih izkusnjah*«
 - i. **Znak &**
 - j. Besede, kjer zapis ni tuja beseda, ampak samo s tujo črko zapisana slovenska beseda.

- i. »full pester promet« ali »Žal . Še huje je to , da raste eksponentno . So pred dnevi krožili neki grafi . Poglejte , blizu te krivulje«

5 Zapis

Cilj: Vsak kodni preklap je označen kot povsem podomačen, delno podomačen ali nepodomačen.

Smernice:

1. Pri presoji zapisa upoštevamo osnovno obliko oz. jedro besede, ne morebitnih obrazil in končnic. Primeri, kot so »downloaddala« ali »trailerji«, so torej označeni kot nepodomačeni.
- a. Kot nepodomačene označimo tudi zapise, ki so zapisani nestandardno glede na izvirni jezik, a sprememba ne odraža »poslovenjenja«, temveč neformalni slog pisanja: *waaaaaaay better, pls, fcukers, I luv bieber, ima se, moze se*
2. (Delno) podomačen zapis označimo pri besedah, kjer je opazna sprememba izvirnega zapisa, sicer jih označimo kot nepodomačene (npr. pri besedah BSH). Delno prilagojen je zapis v naslednji primerih:
 - a. besede z izpuščenim končnim e-jem: *simpl, nop*
 - b. primeri, kjer kljub podomačitvi ostanejo tuji elementi: *gutenacht, Wuuhuu*
 - c. besede, ki so prilagojene (npr. izgubijo dvojni soglasnik), a ne odražajo v celoti »slovenske« izgovorjave: *Haštag, helo, Uber alles, jup*
 - d. »mešani primeri«, kjer je del besedne zveze povsem prilagojen, del pa je zapisan v skladu z izvirnikom (po logiki PP+NP=DP): *ofišl app, velkom back, sori to dissapoint you, dbest trailer*

6 Morfologija

Cilj: Kodne preklope označimo glede na to, ali je tujejezična beseda uporabljena z dodano slovensko končnico (spreganje/sklanjanje), dodanim obrazilom (besedotvorje) ali z obema.

Smernice:

1. Večbesedne zveze (gl. 3. razdelek) obravnavamo kot celoto, in sicer dodamo ustrezno oznako glede na jedro besedne zveze. V primeru »old boysov« je denimo razvidna končnica.
2. Primeri za ROK (razvidna končnica in obrazilo):
 - a. *ofarbane, cheerleaderca, potegal, zalaufi, Prešaltu, Skenslaš, douchebagovskega*
 - b. Primere, kot sta »laufer« in »marketingar« v imenovalniku obravnavamo kot primere z ničto končnico, zato jih označimo z ROK

7 Besedna vrsta

Cilj: Vsak kodni preklop ima pripisano ustrezno oznako za besedno vrsto.

Smernice:

1. Smiselno upoštevamo Osnovne smernice (Oblikoskladenjsko označevanje) in napotke v 3. razdelku teh Smernic.
 - a. Cele stavke ali daljše besedne zveze, ki jim ne moremo jasno določiti besedne vrste, označimo z »neuvrščeno« oz. N.
 - i. »*It was close.*« (MS) ali »*a zdej še tviterji težijo z add your birthday*« (ZS)
 - ii. Enako označimo primere, kot so *feeling comfortable, installing updates, watching basketball*
 - b. Samostalniške zveze z desnim ali levim prilastkom označimo s S: *selfie queen, personal assistant*
 - c. Predložne zveze iz predloga in samostalnika označimo z D: *for your information; down under*
 - d. Pridevniške zveze z desnim ali levim prilastkom označimo s P: *free like a birdy ali waaaay better*
2. Pri problematičnih problemih, kjer pride do »mešanja kodov«, označimo samo tujo besedo in njeno besedno vrsto, ne glede na to, da se dejansko pojavi v predložni zvezi s slovenskim predlogom: na *repeat* (S), v *offline* (R), na *easy* (P)
3. Kot medmete (M) glede na njihovo funkcijo v besedilu označujemo tudi naslednje:
 - a. *please, sori, ofak, holy fuck, fakof, fuck yeah, damn, bummer*
 - b. Klepetalniške krajšave: lol, rofl, OMFG, BTW, WTF, FFS.
4. Kot členke (L) označujemo npr.:
 - a. da (*jp, jap, jep, jup*), ne (*nope, nah*) = členek (L);
 - b. **Izjema:** primere, kot sta »*jesssss*« ali »*Yes!!*« označimo kot medmete (upoštevamo funkcijo).
 - c. Besedi *kao* in *val(j)da*.

8 Oznake za označevanje in njihov pomen

Plast	Oznake	Opis
1. Jezik (odprta kategorija)	EN DE HBS SP LA AR FR IT PT	- angleščina - nemščina - srbohrvaščina - španščina - latinščina - arabščina - francoščina - italijanščina - portugalščina
2. Vrsta preklopa	MS ZS	- medstavčni preklon - znotrajstavčni preklon
3. Zapis	NP DP PP	- nepodomačen - delno podomačen - povsem podomačen
4. Morfologija	NR RK RO ROK	- ni razvidna (brez slovenske končnice/obrazila) - razvidna iz končnice (spreganje/sklanjanje) - razvidna iz obrazila (besedotvorje) - razvidna iz obrazila in končnice
5. Besedna vrsta	G S P R O M L N D V K Z	- glagolska zveza - samostalniška zveza - pridevniška zveza - prislovna zveza - okrajšava - medmet - členek - neuvrščeno - predložna zveza - veznik - števnik - zaimék
6. Hešteg	hashtag	To oznako za razliko od ostalih označimo samo pri heštegih.

7.2 Priloga B: Dodatni primeri zanimivih kodnih preklpov

Seznam znotrajstavčnih preklpov iz manj pogostih jezikov (razen *ala* so se vsi pojavili enkrat)

<u>Št.</u>	<u>Besedilo preklpa</u>	<u>Jezik</u>	<u>Št.</u>	<u>Besedilo preklpa</u>	<u>Jezik</u>
1.	bon voyage	fr	13.	Magna Cum Laude	la
2.	ala	fr	14.	deux ex machina	la
3.	ala	fr	15.	ad hominem	la
4.	qu'est-ce que c'est	fr	16.	obrigada	pt
5.	žnsekua	fr	17.	O-C-U-P-A-D-O	sp
6.	faux pas-jev	fr	18.	el clasica	sp
7.	da capo	it	19.	no bueno	sp
8.	Prosecco	it	20.	buenas noches	sp
9.	Carta Magnum libertatum	la	21.	per favor	sp
10.	cyptotrama aspratum	la	22.	bachate	sp
11.	panus lecomtei	la	23.	#bachata	sp
12.	Nomen est omen	la			

Seznam preklpov, dolgih vsaj osem besed (vsi so se pojavili enkrat)

<u>Št.</u>	<u>Besedilo preklpa</u>	<u>Jezik</u>
1.	Ej,neka se piše, pa makar i dobro	hbs
2.	damn internet connection, cant even see my boo	en
3.	not important, relevant, TBD, highly relevant, not to be deleted at any cause	en
4.	And no, the talking part ... not gonna happen	en
5.	2/2 and you go look at yoour wedding pics	en
6.	fierce like Alexander Avalon Duke the 1st.	en
7.	I dont even know who i am anymore. #sigh	en
8.	yes, I do have a thing for skinny british dudes	en
9.	i thought some Joe guy killed Bruce's parents	en
10.	And i love Batman, he is my fave	en
11.	three bridges connect a cute little medieval town with Slovenia's big capital	en
12.	Hoy en la casa natal de Hugo Wolf en Slovenj Gradec, Eslovenia	sp
13.	#scarf #pullover #masturbatortrenchcoat #cute #funky #stylish #mystyle #newhaircut #asshole	en
14.	good girls go to heaven, bad girls go everywhere else	en
15.	Rihanna goes for acoustic sound with Paul McCartney	en
16.	everything is ready for home teater...movie time...and Verve	en
17.	I'm sure God has a plan for them	en
18.	Last minute shopping at @HamburgAirport. ;) Next stop: home sweet home	en
19.	You read my mind, baby, you are good	en
20.	once you go anal you can never go back	en
21.	you are extremely close to unfollow... extremely close	en
22.	Omg, FB is down..Mars has had running water	en
23.	born in a broken-up poor family without any help	en
24.	You gotta do a tribute! It's a sign	en

Seznam štiribesednih preklpov z oznako HBS (vsi so se pojavili enkrat)

1. I tako sto puta
2. Ima se, moze se
3. bilo kuda, Grahek svuda
4. isto sranje drugo pakovanje
5. Pa ko je ovaj
6. E, nečeš samo tako
7. nek se vidi raskos
8. samo da se priča

7.3 Priloga C: Seznami kodnih preklapov po besednih vrstah

Seznam vseh medmetov (P = število pojavitev)

<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>P</u>	<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>P</u>
1.	#cheers	en	1	54.	LoooL	en	2
2.	#lol	en	3	55.	loool	en	2
3.	#nomnom	en	1	56.	LOOOOOL	en	1
4.	#tenks	en	1	57.	majkemi	hbs	1
5.	#YOLO	en	1	58.	max	en	1
6.	aawww	en	1	59.	Meh	en	3
7.	ahm	en	1	60.	my ass	en	1
8.	awww	en	1	61.	my bad	en	4
9.	awwww	en	2	62.	NOM NOM NOM	en	1
10.	Awwww	en	1	63.	NOOO	en	1
11.	bee-do bee-do bee-do	en	1	64.	obrigada	pt	1
12.	Blah	en	1	65.	ofak	en	1
13.	bon voyage	fr	1	66.	Ofakingsit	en	1
14.	Boom	en	1	67.	oh	en	1
15.	Booo	en	1	68.	Oh c'mon	en	1
16.	btw	en	11	69.	oh heeeelll	en	1
17.	buenas noches	sp	1	70.	oh joy	en	1
18.	Bullshit	en	1	71.	Oh well	en	1
19.	bummer	en	1	72.	oh yeaah	en	1
20.	cheers	en	3	73.	omfg	en	1
21.	damn	en	1	74.	omg	en	8
22.	damn right	en	1	75.	omg omg omg	en	1
23.	damnit	en	1	76.	ouje	en	1
24.	dbest	en	2	77.	oujeah	en	1
25.	džiiiizs	en	1	78.	Pffft	en	1
26.	đizs	en	1	79.	please	en	1
27.	ehm	en	1	80.	pls	en	2
28.	ej es ej pi	en	1	81.	plz	en	1
29.	erm	en	1	82.	plz thx bai	en	1
30.	Fakof	en	1	83.	Pritiplis	en	1
31.	ftw	en	2	84.	RAWR	en	1
32.	FU	en	1	85.	rofl	en	1
33.	fuck	en	1	86.	sorči	en	1
34.	FUCK YEAH	en	1	87.	sori	en	4
35.	gad damit	en	1	88.	sorry	en	3
36.	gag	en	1	89.	sory	en	1
37.	good luck	en	2	90.	tenks	en	3
38.	gotcha	en	1	91.	thx	en	4
39.	hajfajf	en	1	92.	tnx	en	5
40.	Hello	en	1	93.	ugh	en	1
41.	Helo	en	1	94.	uhm	en	1
42.	Hey	en	1	95.	w00t	en	1
43.	Holly sweet f***	en	1	96.	wau	en	1
44.	holy fuck	en	1	97.	waw	en	1
45.	Holy shit	en	1	98.	well	en	3
46.	huh	en	1	99.	Wiiiiiii	en	1
47.	imho	en	2	100.	woohoo	en	2
48.	Jessssssss	en	1	101.	Woohooo	en	1
49.	Khm	en	4	102.	wooooot	en	1
50.	Lawl	en	1	103.	Woop woop	en	1
51.	lol	en	54	104.	wtf	en	3
52.	lolwut	en	1	105.	wUUUUUU	en	1
53.	LooL	en	1	106.	Yes	en	1

Seznam angleških členkov

<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Št. pojavitev</u>
1.	jap	4
2.	jep	2
3.	jp	14
4.	jp jp	1
5.	jp, jp	1
6.	Jpjp	1
7.	jup	4
8.	nah	2
9.	nop	6
10.	nope	6
11.	noup	2
12.	right	1
13.	the	1
14.	yes	2

Seznam nemških glagolov in pridevnikov (BV = besedna vrsta)

<u>Št.</u>	<u>Besedilo</u> <u>preklopa</u>	<u>Ort.</u>	<u>Morf.</u>	<u>BV</u>	<u>Št.</u>	<u>Besedilo</u> <u>preklopa</u>	<u>Ort.</u>	<u>Morf.</u>	<u>BV</u>
1.	blinka, blinka	NP	RK	G	14.	šaltam	PP	RK	G
2.	fila	PP	RK	G	15.	štukat	PP	RK	G
3.	fila	PP	RK	G	16.	zalaufi	NP	ROK	G
4.	hajcajo	PP	RK	G	17.	frej	PP	NR	P
5.	laufa	NP	RK	G	18.	frej	PP	NR	P
6.	laufa	NP	RK	G	19.	ofarbane	PP	ROK	P
7.	laufas	NP	RK	G	20.	pofarban	PP	ROK	P
8.	laufaš	NP	RK	G	21.	prheftan	PP	ROK	P
9.	laufat	NP	RK	G	22.	prštiman	PP	ROK	P
10.	laufat	NP	RK	G	23.	sešraufana	PP	ROK	P
11.	mus	PP	NR	G	24.	švoh	PP	NR	P
12.	prešaltu	PP	ROK	G	25.	zaštofan	PP	ROK	P
13.	rufni	PP	RK	G					

7.4 Priloga D: Seznami kodnih preklapov glede na integracijo

Seznam vseh ortografsko povsem podomačenih kodnih preklapov (BV = besedna vrsta, P = število ponovitev)

Št.	Besedilo preklopa	Jezik	BV	P	Št.	Besedilo preklopa	Jezik	BV	P
1.	#hepi	en	P	1	48.	lajkarjev	en	S	1
2.	#kjut	en	P	1	49.	lajke	en	S	1
3.	afkors	en	R	1	50.	ledr	de	S	1
4.	Aj dont anderstend	en	N	1	51.	mus	de	G	1
5.	ala	fr	D	1	52.	najs	en	P	2
6.	bejbi fejs	en	S	1	53.	nigerkam	en	S	1
7.	Bring on zi žrmans	en	N	1	54.	noup	en	L	2
8.	bulšit	en	S	1	55.	ober-haupt-šturm firerja	de	S	1
9.	dbest	en	M	3	56.	ofak	en	M	1
10.	drajv	en	S	1	57.	Ofakingsit	en	M	1
11.	džiiiizs	en	M	1	58.	ofarbane	de	P	1
12.	ej es ej pi	en	M	1	59.	oldskul marketingar	en	S	1
13.	fakin	en	P	1	60.	osom	en	P	1
14.	faking	en	R	1	61.	ouje	en	M	1
15.	faking	en	P	1	62.	pajsa	de	S	1
16.	faking hot	en	P	1	63.	petrolfrike	en	S	1
17.	Fakof	en	M	1	64.	pofarban	de	P	1
18.	favorital	en	G	1	65.	potegal	en	G	1
19.	fila	de	G	2	66.	potegat	en	G	1
20.	filing	en	S	2	67.	prešaltu	de	G	1
21.	frej	de	R	3	68.	prheftan	de	P	1
22.	frendi	en	S	1	69.	prštiman	de	P	1
23.	hajcajo	de	G	1	70.	red velvet kapkejke	en	S	1
24.	hajfajf	en	M	1	71.	rokenfakingrol	en	S	1
25.	hajpa	en	S	1	72.	rufni	de	G	1
26.	hendlom	en	S	1	73.	sešraufana	de	P	1
27.	hengava	en	G	1	74.	simpl	en	P	2
28.	hepi	en	P	2	75.	Skenslaš	en	G	1
29.	in da ekšn	en	D	1	76.	sorči	en	M	1
30.	invajt	en	S	1	77.	Sori	en	M	4
31.	jap	en	L	4	78.	šaltam	de	G	1
32.	jep	en	L	2	79.	šejkam	en	G	1
33.	Jessssssssss	en	M	1	80.	ŠERAJ	en	G	1
34.	jp	en	L	14	81.	šeral	en	G	1
35.	jp jp	en	L	1	82.	šnelfukarce	de	S	1
36.	jp. jp	en	L	1	83.	špukeca	de	S	1
37.	Jpjp	en	L	1	84.	štukat	de	G	1
38.	kenslat	en	G	1	85.	švoh	de	P	1
39.	kenslata	en	G	1	86.	Tenks	en	M	3
40.	kjut	en	P	4	87.	tvitmit	en	S	1
41.	kjuut	en	P	1	88.	u iber	de	R	1
42.	komp	en	S	1	89.	uibr	de	R	1
43.	kompa	en	S	1	90.	updejtala	en	G	1
44.	kompjutr	en	S	1	91.	zagejmamo	en	G	1
45.	kompu	en	S	1	92.	zaštofan	de	P	1
46.	Lajkajte	en	G	1	93.	zizntregerje	de	S	1
47.	lajkam	en	G	1	94.	žnsekua	fr	N	1

Seznam vseh morfološko integriranih kodnih preklpov (BV = besedna vrsta, P = število ponovitev)

Št.	Besedilo preklopa	Jezik	BV	Morf.	P	Št.	Besedilo preklopa	Jezik	BV	Morf.	P
				<u>int.</u>						<u>int.</u>	
1.	accounta	en	S	RK	1	65.	monster truck-e	en	S	RK	1
2.	bachate	sp	S	RK	1	66.	mtb-jkah	en	S	RK	1
3.	backupa	en	S	RK	1	67.	ne-followa	en	S	ROK	1
4.	bad karme	en	S	RK	1	68.	nigerkam	en	S	ROK	1
5.	beati	en	S	RK	1	69.	odfollowal	en	G	ROK	1
6.	blinka, blinka	de	G	RK	1	70.	ofarbane	de	P	ROK	1
7.	bockat	hbs	G	RK	1	71.	old boysov	en	S	RK	1
8.	bra dizajnerjem	en	S	RK	1	72.	oldskul marketingar	en	S	ROK	1
9.	cheap publicity stunti	en	S	RK	1	73.	ownajo	en	G	RK	1
10.	cheerleaderca	en	S	ROK	1	74.	pajsa	de	S	RK	1
11.	cheese emojiije	en	S	RK	1	75.	patike	hbs	S	RK	1
12.	cookije	en	S	RK	1	76.	pet voyagerjem	en	S	RK	1
13.	douchebagovskega	en	P	ROK	1	77.	petrolfriike	en	S	RK	1
14.	downloadala	en	G	RK	1	78.	pofarban	de	P	ROK	1
15.	dropne	en	G	RK	1	79.	potegal	en	G	ROK	1
16.	efforta	en	S	RK	1	80.	potegat	en	G	ROK	1
17.	el clasica	sp	S	RK	1	81.	predivna	hbs	P	RK	1
18.	everyday playlista	en	S	RK	1	82.	prešaltu	de	G	ROK	1
19.	faux pas-jev	fr	S	RK	1	83.	prheftan	de	P	ROK	1
20.	favorital	en	G	RK	1	84.	prolazu	hbs	S	RK	1
21.	fila	de	G	RK	2	85.	proper roadsteraša	en	S	ROK	1
22.	flatku	en	S	ROK	1	86.	prštiman	de	P	ROK	1
23.	flatu	en	S	RK	1	87.	red velvet kapkejke	en	S	RK	1
24.	followerjev	en	S	RK	2	88.	rentala	en	G	RK	1
25.	followi	en	S	RK	1	89.	rentov	en	G	RK	1
26.	fotoshopat	en	G	RK	1	90.	retardih	en	S	RK	1
27.	frendi	en	S	RK	1	91.	retwital	en	G	RK	1
28.	gangbange	en	S	RK	1	92.	rufni	de	G	RK	1
29.	Gay pride-a	en	S	RK	1	93.	se šunja	hbs	G	RK	1
30.	gimpom	en	S	RK	1	94.	sešraufana	de	P	ROK	1
31.	guramo	hbs	G	RK	1	95.	sharinga	en	S	RK	1
32.	hajcajo	de	G	RK	1	96.	six packom	en	S	RK	1
33.	hajpa	en	S	RK	1	97.	Skenslaš	en	G	ROK	1
34.	handlal	en	G	RK	1	98.	stat/dyn routingom	en	S	RK	1
35.	handlov	en	S	RK	1	99.	streama	en	S	RK	1
36.	hashtaga	en	S	RK	1	100.	šaltam	de	G	RK	1
37.	hashtagov	en	S	RK	1	101.	šejkam	en	G	RK	1
38.	hendlom	en	S	RK	1	102.	ŠERAJ	en	G	RK	1
39.	hengava	en	G	RK	1	103.	šeral	en	G	RK	1
40.	hostama	en	G	RK	1	104.	šneľfukarce	de	S	ROK	1
41.	i dont give a fuck' attitudea	en	S	RK	1	105.	špukeca	de	S	ROK	1
42.	ITju	en	S	RK	1	106.	štukat	de	G	RK	1
43.	izljubimo i ispričamo	hbs	N	RK	1	107.	taxija	en	S	RK	1
44.	jača strana	hbs	S	RK	1	108.	taxijih	en	S	RK	1
45.	jinxat	en	G	RK	1	109.	temp file	en	S	RK	1
46.	jolokio	en	S	RK	1	110.	then assholeu	en	S	RK	1
47.	kenslat	en	G	RK	1	111.	torrente	en	S	RK	1
48.	kenslata	en	G	RK	1	112.	touch screenu	en	S	RK	1
49.	kiwiji	en	S	RK	1	113.	trailerji	en	S	RK	1
50.	kompa	en	S	RK	1	114.	trinidadada	en	S	RK	1
51.	kompu	en	S	RK	1	115.	unfriendanja	en	S	ROK	1
52.	kutiji	hbs	S	RK	1	116.	unspecified errorjem	en	S	RK	1
53.	lagane	hbs	P	RK	1	117.	updejtala	en	G	RK	1
54.	lagerju	de	S	RK	1	118.	upgrade-ana	en	P	ROK	1
55.	Lajkajte	en	G	RK	1	119.	usb hubi	en	S	RK	1

<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>BV</u>	<u>Morf. int.</u>	<u>P</u>	<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>BV</u>	<u>Morf. int.</u>	<u>P</u>
56.	lajkam	en	G	RK	1	120.	uštipce	hbs	S	RK	1
57.	lajkarjev	en	S	ROK	1	121.	webu	en	S	RK	1
58.	lajke	en	S	RK	1	122.	wheela	en	S	RK	1
59.	laufa	de	G	RK	2	123.	zagejmamo	en	G	ROK	1
60.	laufanje	de	S	RK	1	124.	zalaufi	de	G	ROK	1
61.	laufas	de	G	RK	1	125.	zaštofan	de	P	ROK	1
62.	laufaš	de	G	RK	1	126.	Zblendala	en	G	ROK	1
63.	laufat	de	G	RK	2	127.	zizntregerje	de	S	RK	1
64.	laufer	de	S	ROK	1						

Seznam vseh ortografsko in morfološko integriranih preklpov

<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>BV</u>	<u>Ort. int.</u>	<u>Morf. int.</u>	<u>Št.</u>	<u>Besedilo preklopa</u>	<u>Jezik</u>	<u>BV</u>	<u>Ort. int.</u>	<u>Morf. int.</u>
1.	fila	de	G	PP	RK	27.	prheftan	de	P	PP	ROK
2.	fila	de	G	PP	RK	28.	prštiman	de	P	PP	ROK
3.	hajcajo	de	G	PP	RK	29.	sešraufana	de	P	PP	ROK
4.	prešaltu	de	G	PP	ROK	30.	zaštofan	de	P	PP	ROK
5.	rufni	de	G	PP	RK	31.	pajsa	de	S	PP	RK
6.	šaltam	de	G	PP	RK	32.	šnělfukarce	de	S	PP	ROK
7.	štukat	de	G	PP	RK	33.	špukeca	de	S	PP	ROK
8.	favorital	en	G	PP	RK	34.	zizntregerje	de	S	PP	RK
9.	fotoshopat	en	G	DP	RK	35.	bra dizajnerjem	en	S	DP	RK
10.	handlal	en	G	DP	RK	36.	cookije	en	S	DP	RK
11.	hengava	en	G	PP	RK	37.	followi	en	S	DP	RK
12.	kenslat	en	G	PP	RK	38.	frendi	en	S	PP	RK
13.	kenslata	en	G	PP	RK	39.	hajpa	en	S	PP	RK
14.	Lajkajte	en	G	PP	RK	40.	handlov	en	S	DP	RK
15.	lajkam	en	G	PP	RK	41.	hendlom	en	S	PP	RK
16.	potegal	en	G	PP	ROK	42.	kompa	en	S	PP	RK
17.	potegat	en	G	PP	ROK	43.	kompu	en	S	PP	RK
18.	retwital	en	G	DP	RK	44.	lajkarjev	en	S	PP	ROK
19.	Skenslaš	en	G	PP	ROK	45.	lajke	en	S	PP	RK
20.	šejkam	en	G	PP	RK	46.	nigerkam	en	S	PP	ROK
21.	ŠERAJ	en	G	PP	RK	47.	oldskul				
22.	šeral	en	G	PP	RK		marketingar	en	S	PP	ROK
						48.	petrolfrike	en	S	PP	RK
						49.	red velvet				
23.	updejtala	en	G	PP	RK		kapkejke	en	S	PP	RK
24.	zagejmamo	en	G	PP	ROK	50.	temp file	en	S	DP	RK
25.	ofarbane	de	P	PP	ROK						
26.	pofarban	de	P	PP	ROK						

7.5 Priloga E: Seznam najpogostejših besed, ki so del kodnih preklapov

Seznam besed, ki so se najpogosteje ponovile v kodnih preklapi (P = število ponovitev). Za večjo nazornost smo združili število ponovitev za različne oblike zapisov (npr. »LOL«, »lol«, »LoL«, »loool« ipd. se štejejo kot ena beseda, ki se v vseh oblikah skupaj ponovi 63-krat).

Št.	Beseda	P	Št.	Beseda	P	Št.	Beseda	P	Št.	Beseda	P
1.	I, I'm	77	52.	it's	8	103.	awwww, ipd.	4	154.	break	3
2.	LOL, loool, ipd.	63	53.	more	8	104.	Back	4	155.	bring	3
3.	you, youll, you're	40	54.	Well	8	105.	Big	4	156.	close	3
4.	it	31	55.	Yes	8	106.	breath	4	157.	coffee, coffee	3
5.	jep, jap, jp	30	56.	kjut, kjuut	7	107.	but	4	158.	cookie, cookije	3
6.	a	30	57.	Also	7	108.	Can't	4	159.	damit, damnit	3
7.	RT	30	58.	guess	7	109.	cheers	4	160.	dance	3
8.	the	28	59.	nice	7	110.	clap	4	161.	eyes	3
9.	to	27	60.	se	7	111.	crap, C-R-A-P	4	162.	Fast	3
10.	my	26	61.	that	7	112.	damn	4	163.	fav	3
11.	of	25	62.	this	7	113.	dbest	4	164.	fire	3
12.	fuck, fak	23	63.	goes	6	114.	ever	4	165.	frej	3
13.	good	23	64.	work, working	6	115.	family	4	166.	Give	3
14.	thanks, tnx, thx	19	65.	again	6	116.	feel	4	167.	has	3
15.	and	19	66.	can	6	117.	feeling	4	168.	hate	3
16.	no	19	67.	da	6	118.	friend, friends	4	169.	help	3
17.	IN	18	68.	day, days	6	119.	great	4	170.	hint	3
18.	for	17	69.	laufa, laufaš	6	120.	guys, guy	4	171.	idea	3
19.	know, knows	16	70.	makes	6	121.	have, having	4	172.	Ima	3
20.	on	15	71.	So	6	122.	home	4	173.	je	3
21.	nop, nope, noup	14	72.	wtf	6	123.	if	4	174.	joke	3
22.	happy, hepi	14	73.	they, They're	5	124.	Khm	4	175.	Keep	3
23.	me	14	74.	about	5	125.	kind	4	176.	lost	3
24.	all	13	75.	am	5	126.	late	4	177.	Meh	3
25.	be	13	76.	any	5	127.	look, Looks	4	178.	mission	3
26.	do, does	13	77.	Best	5	128.	luck	4	179.	morning	3
27.	GO	13	78.	by	5	129.	man	4	180.	news	3
28.	much	13	79.	cc	5	130.	MT	4	181.	NOM	3
29.	not	13	80.	cool	5	131.	never	4	182.	Non	3
30.	omg	13	81.	deal	5	132.	next	4	183.	people	3
31.	vs, vs., v, v.	13	82.	did	5	133.	Now	4	184.	photo	3
32.	btw	12	83.	down	5	134.	old	4	185.	read(ing)	3
33.	your, yoor	11	84.	exactly	5	135.	or, ooor	4	186.	ready	3
34.	is	11	85.	flat, flatku, flatu	5	136.	Pretty	4	187.	red	3
35.	time	11	86.	fun	5	137.	said	4	188.	repeat	3
36.	too	11	87.	get	5	138.	selfie	4	189.	right	3
37.	sorry, sori	10	88.	girl, girls	5	139.	Slow	4	190.	Run(ing)	3
38.	don't, dont	10	89.	he	5	140.	stop	4	191.	s*x	3
39.	JUST	10	90.	hot	5	141.	up	4	192.	simpl	3
40.	kao	9	91.	how	5	142.	us	4	193.	Social	3
41.	oh	9	92.	little	5	143.	welcome	4	194.	Some	3
42.	shit	9	93.	off	5	144.	what, Whaaat	4	195.	stream	3
43.	U	9	94.	one	5	145.	Why	4	196.	take	3
44.	with	9	95.	respect	5	146.	woohoo, wuuhuu	4	197.	taxi	3
45.	like, likes	8	96.	see	5	147.	hashtag, haštag	3	198.	thought	3
46.	love, luv	8	97.	sense	5	148.	komp, kompa,	3	199.	trailer	3
47.	pls, plz, please	8	98.	There	5	149.	ass	3	200.	update	3
48.	are	8	99.	Valda, valjda	5	150.	awesome	3	201.	way	3
49.	at	8	100.	Haha, hah	4	151.	badass	3	202.	what's	3
50.	bad	8	101.	them	4	152.	bee-do	3	203.	will	3
51.	free	8	102.	As	4	153.	bit	3	204.	true	3

8 Izjava o avtorstvu

Izjavljam, da je magistrsko delo v celoti moje avtorsko delo ter da so uporabljeni viri in literatura navedeni v skladu z mednarodnimi standardi in veljavno zakonodajo.

Ljubljana, 6. 9. 2017

Špela Reher